引　言

在当代社会中，信息和信息科学这两个名词已为人们所熟知。它们的重要作用也正在被人们所接受。但究竟什么是信息？什么是信息科学？信息到底能否度量？怎样度量？对这些问题人们往往不十分清楚，从最广义的情形来说，这些问题涉及到信息的哲学解释与哲学地位。在本书中，我们不打算讨论这种一般性的问题，而是试图介绍几种关于信息度量的比较成熟的定义形式，以及它们的产生过程、性质与应用。我们希望，通过这些讨论可以帮助人们理解有关信息、信息的度量与信息科学的内容与意义，以及它们的概念特征与应用范围。

由于信息概念的广泛性，试图对信息的一般形式进行度量是十分困难的。本世纪20年代，奈奎斯特(H.Nyquist)与哈特莱(L.Hartley)就已指出了信息度量与通信理论的关系，以及它们与概率、对数函数的联系。1948年在仙农(C.E.Shannon)的著名论文《通信的数学理论》中，对信息的度量、通信中的编码问题给出了一系列确切的分析与论述，这些论述在以后的理论发展中得到了充分的证实与应用。因此，人们往往把信息论的产生与仙农的工作相联系，有关仙农所引进的信息度量被称为仙农熵，而相应的编码理论被称为仙农信息论。

VISIT…

LANZAROTE Caliente.COM

仙农熵在信息度量中的成功不仅在于它具有明确的内在含意与严格的数学表达，而且它能确切地刻划出通信中的一系列本质特征。如由仙农熵确定的信号体积等概念就是信源在通信中的一种本质特征；又如由它派生的交互信息与信道容量反映了通信信道的基本特征。由仙农熵所确定的信息度量单位—“比特”已是通信理论与计算机科学中的一个基本单位。

由于信息概念的广泛性，关于信息度量的推广与一般形式的表达问题的探讨一直是人们密切关心的一个重大问题，任何新的度量形式出现都会导致新的信息科学的学科分支出现。除了由仙农熵及其派生或推广的信息度量之外，还可有与仙农熵有完全不同出发点的其它信息度量。例如算法信息论中的柯尔莫各洛夫(Kolmogorov) 复杂性等。因此，我们对信息度量的理解不仅要从它们的引进来源来理解，更重要的是要从这些量的应用特征来理解，这样就可使我们更好地掌握与应用信息度量的工具。

本书的主要目的就是希望介绍几种主要的信息量及它们所涉一些分支学科，尤其对仙农熵及其有关的信源、信道编码理论作较为完整的叙述。为了适合读者的不同要求，我们对比较专门的内容章、节用星号注明，对此可以省略不读，而不影响对全书基本内容的理解。另外，本书的参考文献将按专题列出，有兴趣的读者可再深入学习了解。

具有初步微分学与概率论知识的读者均可阅读本书。一些微分学与概率论中的基本知识与名词概念，如什么是集合，集合之间的相互关系与交、并、差、积运算，集合之间的映射(或变换)等，又如对数、指数函数，导数与极值，最大、最小值问题，及概率论中的随机试验，随机事件，随机变量，概率，概率分布，概率密度，均值与方差等等，对这些概念名词我们

不再一一解释，希望读者参考有关知识材料。

作者感谢中国数学会传播委员会及湖南教育出版社对本书写作出版的支持与帮助。

沈世镒

1991年12月25日

有关记号

$\boldsymbol{N}=\{1, 2, \cdots, n\}$，$\boldsymbol{M}=\{1, 2, \cdots, m\}$，

$\boldsymbol{M}'=\{1, 2, \cdots, m'\}$：部分自然数集合。

$\boldsymbol{\Lambda}=\{0, \pm1, \pm2, \cdots\}$，

$\boldsymbol{\Lambda}_+=\{0, 1, 2, \cdots\}$：整数与非负整数集合。

$\boldsymbol{GF}(q)=\{0, 1, \cdots, q-1\}$：$q$-值的有限域。

$\boldsymbol{A}$，$\boldsymbol{B}$，$\boldsymbol{X}$，$\boldsymbol{Y}$，…，$\boldsymbol{U}$，$\boldsymbol{V}$，$\boldsymbol{W}$：集合记号，或称为字母表。

$\boldsymbol{X}\times\boldsymbol{Y}$：$\boldsymbol{X}$，$\boldsymbol{Y}$的乘积空间。

a，b，x，y：分别为集合$\boldsymbol{A}$，$\boldsymbol{B}$，$\boldsymbol{X}$，$\boldsymbol{Y}$中的元素，或称为字母。

X，Y，Z：分别取值于$\boldsymbol{X}$，$\boldsymbol{Y}$，$\boldsymbol{Z}$中的随机变量。

$p(x)=P_r\{X=x\}$：随机变量X取值为x的概率。

$q(y)=P_r\{Y=y\}$：随机变量Y取值为y的概率。

$p(x, y)=P_r\{X=x, Y=y\}$：随机变量(X, Y)取值为(x, y)的概率。

$p(x/y)=P_r\{X=x/Y=y\}$：随机变量X在Y取值为y的条件下取值为x的条件概率。

$q(y/x)=P_r\{Y=y/X=x\}$：随机变量Y在X取值为x的条件下取值为y的条件概率。

$p_i=P_r\{X=i\}$：随机变量X取值为i的概率。

$q_j=P_r\{Y=j\}$：随机变量Y取值为j的概率。

$p_{i,j}=P_r\{X=i, Y=j\}$：随机变量(X, Y)取值为(i, j)

的概率。

$p_{i/j}=P_r\{X=i/Y=j\}$：在Y取值为j的条件下，X取值为i的条件概率。

$q_{j/i}=P_r\{Y=j/X=i\}$：在X取值为i的条件下，Y取值为j的条件概率。

$p(\cdot)=(p(x),\ x\in \boldsymbol{X})$或$p^m=(p_1,\ \cdots,\ p_m)$：

在$\boldsymbol{X}$或$\boldsymbol{M}$中取值的离散随机变量X的概率分布。

$q(\cdot)=(q(y),\ y\in \boldsymbol{Y})$或$q^{m'}=(q_1,\ \cdots,q_{m'})$：

在$\boldsymbol{Y}$或$\boldsymbol{M}'$中取值的离散随机变量X的概率分布。

$p(\cdot,\ \cdot)=(p(x,\ y),\ x\in \boldsymbol{X},\ y\in \boldsymbol{Y})$ 或 $p^{m,m'}=(p_{i,j},\ i\in \boldsymbol{N},\ j\in \boldsymbol{N}')$：$\boldsymbol{X}\times \boldsymbol{Y}$或$\boldsymbol{M}\times \boldsymbol{M}'$上的概率分布。

$\boldsymbol{X}^n=\prod_{i=1}^{n}\boldsymbol{X}_i$：$\boldsymbol{X}_1,\ \cdots,\ \boldsymbol{X}_n$的乘积空间，

$\boldsymbol{X}_i=\boldsymbol{X},i=1,\cdots,\ n$。

$x^n=(x_1,\ \cdots,\ x_n)$：$\boldsymbol{X}^n$中的元，或$\boldsymbol{X}$上取值的n－维向量。

$X^n=(X_1,\cdots,\ X_n)$：$\boldsymbol{X}^n$上取值的随机变量，或$\boldsymbol{X}(=\boldsymbol{X}_i)$上取值的$n$－维随机向量。

$p(x^n)=P_r\{X^n=x^n\}$：随机向量X^n取值为x^n的概率。

$\boldsymbol{Y}^n,\ y^n,\ Y^n,\ q(y^{n'})$：与$\boldsymbol{X}^n,\ x^n,\ X^n,\ p(x^n)$类似定义。

$\log(x),\ \ln(x)$：对数函数，分别以2，e为底数。

$\exp_2(x),\ \exp(x)$：指数函数，分别以2，e为底数。

$h[p(x)]=-\log[p(x)]$ 或 $h(p_i)=-\log p_i$：$p(x)$或p_i的熵密度。

$$H(X)=-\sum_{x\in X}p(x)\log p(x)\ 或H(X)=-\sum_{i=1}^{n}p_i\log p_i:$$

随机变量X(分别在$\boldsymbol{X}$或$\boldsymbol{M}$上取值)的熵。

$h[p(x/y)]=-\log[p(x/y)]$ 或 $h(p_{i/j})=-\log p_{i/j}$：

$p(x/y)$或$p_{i/j}$ 的条件熵密度。

$$H(X/Y)=-\sum_{x\in X}\sum_{y\in Y}p(x,\ y)\log p(x/y)$$

或

$$H(X/Y)=-\sum_{i=1}^{m}\sum_{j=1}^{m'}p_{i,j}\ \log p_{i/j}$$ ：随机变量X关于Y的条件熵。

$h[q(y/x)]=-\log[q(y/x)]$ 或 $h(q_{j/i})=-\log q_{j/i}$ ：$q(y/x)$或$q_{j/i}$ 的条件熵密度。

$$H(Y/X)=-\sum_{x\in X}\sum_{y\in Y}p(x,\ y)\log q(y/x)$$

或

$$H(Y/X)=-\sum_{i=1}^{m}\sum_{j=1}^{m'}p_{i,j}\ \log q_{j/i}$$ ：随机变量Y关于X的条件熵。

$$I(X;\ Y)=\sum_{x\in X}\sum_{y\in Y}p(x,\ y)\log\frac{p(x,\ y)}{p(x)\cdot q(y)}$$ 随机变量X，Y的交互信息。

$$i(x;\ y)=\log\frac{p(x,\ y)}{p(x)\cdot q(y)}$$ 随机变量X，Y的交互信息密度。

$f_E(x)$：$\boldsymbol{X}\to\boldsymbol{U}$，从$\boldsymbol{X}$到$\boldsymbol{U}$的编码函数。

$g_D(v)$：$\boldsymbol{V}\to\boldsymbol{Y}$，从$\boldsymbol{V}$到$\boldsymbol{Y}$的译码函数。

$F(x)$，$P(x)$，$Q(x)$：随机变量X的分布函数，这时

$$F(x)=P(x)=Q(x)=P_r\{X<x\}.$$

$f(x)$，$p(x)$，$q(x)$：连续型分布函数的分布密度，这时

$$f(x)=dF(x)/dx,\qquad p(x)=dP(x)/dx,$$

$$q(x)=dQ(x)/dx.$$

目　录

前言(王　元) 1

引言(沈世镒) 3

有关记号 6

第一章　概　论 ……1

§1.1　仙农信息论的产生、发展与应用……1

§1.2　仙农熵的引进与信息度量的研究状况……5

§1.3　信息论与信息科学…… 12

第二章　仙农熵与无噪声信源编码问题 …… 14

§2.1　通信系统概述…… 14

§2.2　信源的不等长信号编码问题…… 22

§2.3　熵功率与信号体积…… 29

第三章　信道容量与信道编码定理 …… 33

§3.1　条件熵与交互信息…… 33

§3.2　信道容量及其性质…… 39

§3.3　无记忆信道的编码定理…… 45

§3.4　纠错码简介…… 54

第四章　信息量的推广与应用 …… 60

§4.1　连续分布的仙农熵与互熵…… 60

§4.2　最大熵与最小互熵原理…… 68

§4.3　广义熵的定义与性质…… 72

第五章　编码理论的发展与应用 …… 75

§5.1 率失真函数与数据压缩编码定理…………………75
§5.2 多用户通信网络概论………………………………79
§5.3 多重信源的编码定理………………………………83
§5.4 多址信道的容量区域与编码定理…………………90
第六章 信息量在信息科学的其它分支中的应用 ……………95
§6.1 信息量在密码学中的应用…………………………95
§6.2 信息量与计算机复杂性理论与分形几何的关系………………………………………………………99
§6.3 信息量在统计理论中的应用……………………103
结束语 ……………………………………………………104
参考文献 …………………………………………………105
索引 ………………………………………………………109

编后记(冯克勤) 111

第一章 概 论

§1.1 仙农信息论的产生、发展与应用

1.信息论的形成与发展

信息的概念是一个十分广泛的概念，有的哲学家甚至把它看作客观世界中有别于物质、能量的第三大要素，并把它看作是推动当今社会文明的主要因素．它的广泛性不仅涉及到人类社会的各个领域，而且在生物世界也都离不开信息的交流．从动物之间的各种动作交流到细胞的遗传生长都有信息的存在与作用。

在早期的人类社会中，人们的信息交流主要在文字、语言乃至动作表示上，当时就有利用各种信号包括符号来传达信息的种种表现(如烽火、鼓乐等等)，但人们大量的利用信息还是在电子通信与计算机技术出现之后．为了提高通信的质量与效率，人们往往从物理与数学两方面出发来进行研究考虑．在物理方面，主要工作是改进通信的物理手段与条件，如采用微波、卫

星、激光等手段使通信方式发生革命性的变化，同时通过对频带与信噪比等指标的改进也可提高通信的数量与质量。而数学的考虑角度则是在信号的设计构造与编、译码的算法上。也就是，在物理设备条件不变的情况下，用数学的方法来改进通信的数量与质量。在一般的通信问题中，数量与质量是两个相互制约且可相互补偿的指标。这样在保证通信质量的前提下，如何传送最多的信息就成为通信理论中的一个基本问题。因此要求我们在数学上解决一系列问题，例如，如何确定通信中信息传递的数量与质量的评估标准，如何在保证通信质量的前提下，提高它的数量问题及相应的计算机的运算实现等问题。

在本世纪20年代，奈奎斯特与哈特莱就提出了解决上述问题的一系列途径，如信息传递的速率与带宽成比例，信息的度量与信号的概率分布、对数函数相联系等等，这些思想为以后仙农信息论的产生打下了基础。

至本世纪40年代，“控制论”的奠基人维纳 (N.Wiener)、美国统计学家费希尔(E.Fisher)与仙农几乎同时提出了信息度量的熵的定义形式。这个事实说明了从不同的学科出发，都会导致信息量这个概念的产生。1948年，仙农的著名论文《通信的数学理论》被认为是信息论产生的奠基性工作，因为该论文不仅对这种信息的度量作出了明确的描述，而且成功地利用了这种信息度量解决了信源、信道的编码问题。人们把这种信息的度量称为仙农熵，而相应的编码理论称为仙农信息论。仙农熵与仙农信息论是本书介绍的重点。

2.仙农信息论的发展与应用

自1948年仙农信息论产生以来，由于电子、通信与计算机技术发展的突飞猛进，信息论的发展也十分迅速。四十多年来，

该理论大体上经历了理论的确立、理论的应用、理论的发展与近代发展四个阶段，从而大大丰富了该理论的内容。我们现在对这四个阶段的情况作一简单的介绍。

理论的确立阶段大体上是从40年代末到60年代初。在这十余年的时间里，大量的工作是对仙农的奠基性论文《通信的数学理论》的理解、解释与分析。虽然仙农的工作是正确、完整的，但是当时能完全读懂它的人不多。这与维纳的《控制论》一书的情形十分相似。因此，在这个阶段中，许多工作就是为仙农论文的严格表述而写。例如，关于仙农熵的公理化条件；关于信源、信道编码定理的数学表达及通信系统的描述与分类等。该阶段基本完成的标志是一系列专著的出现，如苏联科学院院士欣钦(А.Я.Хичин)于1956年，美国著名的信息论专家范恩斯坦(A.Feinstein)，伽拉格尔(R.G.Gallager)分别于1958年与1968年的专著的出版等。这些著作除了把仙农信息论的内容体系给出了完整的叙述之外，在许多方面又有新的发展，如有记忆信道的编码理论、无记忆信道的编码的误差界理论等。

编码理论的应用实现几乎与仙农理论并行发展，但它的产生历史更早些。在电子通信技术产生不久就有莫尔斯(Morse)码与波多(Bodo)码出现，它们把字母、数字与一串点、划、空等电子信号对应，使消息能够通过信号发送，在早期电子通信中广泛使用，但这种码的构造比较简单，不具有自我纠错的能力。

自50年代开始，由于通信技术的发展，要求信号对干扰具有自身的保护能力，这就导致了一系列新的纠错码的产生。这种纠错码在仙农理论的指导下，综合代数与概率、统计的理论与方法，应用群、环、域等工具，对码进行具体构造，并且利用移位寄存器等工具，建立编、译码的快速算法，使这些纠错码的理论能在快速通信中得以实现。当时，著名的信源编码有

仙农-费诺(Shannon-Fano)码等，而信道编码有汉明(R.W.Hamming)码、BCH码、R-S码、戈帕(Goppa)码等，著名的译码算法有卷积码的维特比(Viterbi)译码算法等。这些编码体制已在通信工程中广泛使用，成为当代通信工程中不可缺少的一个组成部分。值得指出的是编码理论的发展方兴未艾，纠错码的应用范围已不再局限在通信工程内，在计算机内部构造，各种光、电、磁产品(如光盘，磁性记忆存储器，家用音视设备等)中都有使用。因此编码技术在日、美、西欧各国已成为一个热门的行业。

70年代，信息论的发展集中在多用户信息论与率失真理论两大方面。多用户信息论是将简单的单用户通信模型发展成多用户的网络通信模型。这在卫星通信系统(通信卫星、电视卫星等)、其它电子系统中有广泛的应用背景。因此多用户信息论在70年代末、80年代初形成高潮，研究的模型有数十种之多。主要信道模型有多址信道、广播信道、有边信息信道、有反馈信道、转播信道等，多用户信源编码问题有无噪声多重信源编码问题、具有率失真度的多重信源编码问题、多重信源与多路信道结合编码问题与多终端编码问题等。

率失真理论又名数据压缩或数据量化的数学理论，它的应用背景是这样的：为了复制一组实用信号(如映、视图象，语音信号，图、文印刷等)，可有一定的误差允许存在。这些误差在一定的范围内不影响人们的视、听效果。如果我们在信号复制过程中能利用这些允许误差，这样就可大大降低信息提取中的数据量。这在工程界称之为数据压缩或数据量化问题。率失真理论解决了失真度与数据压缩量之间的函数关系，给出了在信息处理的允许误差范围内的最小信息提取要求，并在图象、语音及其它信号处理中有大量的应用。率失真理论也是仙农信

息论的一个典型发展，率失真函数的计算就是在仙农熵与交互信息的基础上建立起来的一个信源编码理论指标。

自80年代末期起，信息论、仙农熵在更广泛的意义上得到发展，它的主要进展有仙农熵与算法信息论、分形几何的结合，互熵与统计中的微分流形方法相结合。这些新的分支领域大大丰富了信息论的发展范围，也使我们看到了仙农熵在自然科学中有更广泛的应用与含义。同时，信息论的观点、方法与结果在密码学，量子信号处理，光、磁信息处理等领域中得到应用，这些内容构成了信息科学中的一个重要部分。对这些分支内容我们在以后的讨论中还有详述。

§1.2 仙农熵的引进与信息度量的研究状况

1.仙农熵的引进

仙农熵的基本概念来自随机试验(或随机变量)的不肯定性。关于随机试验(或随机变量)的概念在概率论中已经阐明，因此我们重点解释随机试验的不肯定性。

正如概率论中所叙述的那样，一个随机试验包含有两个因素，即它的试验可能有多种结果出现，且每个结果的出现具有一定的可能性大小或概率（有的书中称为几率）.如果可能出现的全体结果是有限或可数的，那么我们称这个随机试验是离散的。一个离散的随机试验或随机变量可以用$X=[\boldsymbol{X},p(x)]$来表示，其中$\boldsymbol{X}$为随机试验X的全体可能出现的结果，它的元素$x\in\boldsymbol{X}$为随机试验的基本事件，而$p(x)$为在随机试验X中，基本事件

x出现的概率，以下记

$$p(\cdot)=\{p(x):\ x\in\boldsymbol{X}\}$$

为随机变量X的概率分布.如$\boldsymbol{X}=\boldsymbol{M}=\{1,\ 2,\ \cdots,\ m\}$(有时用$\boldsymbol{M}=\{1,\ 2,\ \cdots,\ m\}$),那么相应的随机试验可表示为:

$$X=\begin{pmatrix}1 & 2 & 3 & \cdots & m\\ p_1 & p_2 & p_3, & \cdots, & p_m\end{pmatrix}. \qquad (2.1)$$

一个随机试验的概率分布有比较集中或分散之区别,因此,它们的不肯定性大小也有区别。如我们比较以下的几种概率分布类型:

$$X_1=\begin{pmatrix}0 & 1\\ 0 & 1\end{pmatrix},\quad X_2=\begin{pmatrix}0 & 1\\ 0.5 & 0.5\end{pmatrix},\quad X_3=\begin{pmatrix}0 & 1\\ 0.1 & 0.9\end{pmatrix},$$

$$X_4=\begin{pmatrix}0 & 1 & 2 & \cdots & 9\\ 0.1 & 0.1 & 0.1 & \cdots & 0.1\end{pmatrix}. \qquad (2.1')$$

我们现在可以分析这四种离散随机试验X_1-X_4的不肯定性。因为随机试验X_1是一个决定性的试验，事件1必然发生，而事件0一定不发生，因此它的不肯定性为零。对随机试验X_2与X_3，它们虽然都有不肯定性,但它们的不肯定大小是明显不同的，因为在随机试验X_3中，事件1虽不一定发生，但它发生的可能性很大，同样，事件0虽不一定不发生,但它发生的可能性很小。而随机试验X_2的情形则完全不同,事件0与1发生的可能性完全相同，因此随机试验X_2的不肯定性明显大于X_3。随机试验X_3与X_4的不肯定性也是不同的,因为随机试验X_4出现的结果比X_3的结果多,且都是均匀分布,所以随机试验X_4的不肯定性比随机试验X_3的不肯定性大。由此可知,随机试验的不肯定大小度量确实存在，而且与随机试验的事件空间大小、概率分布的集中程度有关。我们可以对离散随机试验$X=[\boldsymbol{M},\ p_i]$的不肯定性度量规定为一个概率分布(p_i，$i=1,\ 2,\ \cdots,\ m$)的

函数，也就是

$$H(X)=H(p^m)=H(p_1,\ p_2,\ \cdots,\ p_m). \tag{2.2}$$

我们称 $H(X)$ 为离散随机试验 X 的不肯定性度量或不肯定度。以下我们记

$$\boldsymbol{P}_m=\left\{p^m=(p_1,\ \cdots,\ p_m):\ p_i\geqslant 0,\ i=1,\ \cdots,\ m;\ \sum_{i=1}^{m}p_i=1\right\} \tag{2.3}$$

为 $\boldsymbol{M}$ 上的全体概率分布。这时 $H(X)$ 为 $\boldsymbol{P}_m$ 上的函数。为了对不肯定性度量有一个确切的数学表达，我们以下介绍不肯定度的公理化描述。

关于随机试验不肯定度的公理化条件为：

公理1 在贝努里试验中肯定试验的不肯定度为零，也就是 $H(0,\ 1)=0$。

公理2 在贝努里试验中不肯定度 $H(p,\ 1-p)$ 是 $p\in[0,1]$ 的连续函数。

公理3 不肯定度 $H(X)$ 的大小与 $\boldsymbol{X}$ 中事件排列次序无关，也就是不肯定度函数 $H(p_1,\ p_2,\ \cdots,\ p_m)$ 关于变量 $p_1,\ p_2,\cdots,\ p_m$，是对称的。

公理4 不肯定度的大小具有可加性。也就是如果随机试验的某事件内部蕴含不肯定度，那么这些不肯定度在平均概率意义下是可加的。这时它的数学表达式为

$$\begin{aligned}&H(p_1,\ p_2,\ \cdots,\ p_{m-1},\ q_1,\ q_2)\\&=H(p_1,\ p_2,\ \cdots,\ p_{m-1},\ p_m)+p_m\cdot H(q_1/p_m,\ q_2/p_m),\end{aligned} \tag{2.4}$$

其中 $p_m=q_1+q_2$。

这里提到的贝努里试验是一个二值的随机试验，可用 $X=$

$\begin{pmatrix}0 & 1\\ p & 1-p\end{pmatrix}$来表示，这时$H(X)=H(p,\ 1-p)$。

上述不肯定性的公理化条件中，公理1—公理3是比较自然的，对公理4我们举例说明如下：

如要找某年级的某同学，该年级有100名学生分成4个班，每班有25名学生，那么该同学在该年级的不肯定度应是该同学所在班级的不肯定度与他所在班级内的不肯定度之和，这也是可理解的。

定理1 如果$H(p^m)$为$\boldsymbol{P}_m$上的函数且满足公理1—4，那么$H(p^m)$必为以下对数函数的形式：

$$H(p_1,\ \cdots,\ p_m)=-\sum_{i=1}^{m}p_i\cdot\log_a(p_i),\qquad(2.5)$$

其中$\log_a(\cdot)$为以a为底的对数函数。

该定理在许多信息论的著作中均有论述与证明，有兴趣的读者可参阅[5—8]文。由于以上不肯定度量函数$H(p_1,\cdots,p_m)$的形式与热力学中的熵的形式十分相似，因此该不肯定性的度量函数被称为仙农熵。

在工程界，对(2.5)中对数的底a有不同的选择，常用的底取为$a=2$，3，e，10，并由此产生的不同的信息单位分别为“比特”（Bit，信息度量的二进制单位），“铁特”（Tet，信息度量的三进制单位），“奈特”（Nat，信息度量的自然单位），“笛特”（Det，信息度量的十进制单位）。

例1 计算(2.1′)中随机试验X_2的仙农熵。

解：在随机试验X_2中，概率分布为$p_0=p_1=1/2$，因此

$$H(X_2)=-\frac{1}{2}\log\left(\frac{1}{2}\right)-\frac{1}{2}\log\left(\frac{1}{2}\right)=1.$$

因此(2.1′)中的随机试验X_2的不肯定性（或仙农熵）为信息的

度量单位“比特”。这是工程中最常用的信息度量单位。

2.仙农熵的基本性质

为讨论仙农熵的性质，我们先给出若干关于对数函数的性质。

引理1 关于对数函数有以下不等式成立：

(1) 对任何$x>0$，有不等式

$$1-1/x\leqslant \ln x\leqslant x-1 \tag{2.6}$$

成立，且等号成立的充要条件为$x=1$。

(2) $f(x)=\log x$是x定义域中的下凸函数，也就是对任何x_1，$x_2>0$，总有不等式

$$\log\left[\frac{1}{2}(x_1+x_2)\right]\geqslant\frac{1}{2}\left[\log x_1+\log x_2\right] \tag{2.7}$$

成立。

证明：为证(2.6)式，我们设函数$g(x)=x-1-\ln x$，这时

$g'(1)=(1-1/x)|_{x=1}=0, g(1)=0, g''(x)=1/x^2>0$且对任何$x>0$成立，因此$g(x)$在$x=1$处取最小值，且最小值为零，因此$g(x)\geqslant 0$或$\ln x\leqslant x-1$对任何$x>0$成立，且等号成立的充要条件为$x=1$。关于(2.6)的第一个不等式由$\ln(1/x)\leqslant 1/x-1$即得$\ln x\geqslant 1-1/x$成立。

$f(x)=\log x$在$x>0$的定义域中的下凸性由算术、几何平均不等式

$$\frac{1}{2}(x_1+x_2)\geqslant\sqrt{x_1\cdot x_2} \tag{2.7'}$$

即得。引理得证。

引理2 关于算术、几何平均不等式还可推广为以下一般形

式，即对任何两组 p_i，$q_i>0$，$i=1, 2, \cdots, m$，且 $p_1+\cdots+p_m=1$，总有

$$\sum_{i=1}^{m} p_i \cdot \log q_i \leqslant \log\left[\sum_{i=1}^{m} p_i \cdot q_i\right] \tag{2.8}$$

成立，且等号成立的充要条件为$q_1=q_2=\cdots=q_m$。

证明：如记 $Q=\sum\limits_{i=1}^{m} p_i \cdot q_i$，那么由引理1得

$$\sum_{i=1}^{m} p_i \cdot \log q_i - \log Q = \sum_{i=1}^{m} p_i \cdot \log(q_i/Q)$$

$$\leqslant \sum_{i=1}^{m} p_i(q_i/Q-1)=0$$

成立，其中不等式中等号成立的充要条件为$q_1=q_2=\cdots=q_m=Q$。引理得证。

引理3 对引理2中的p_i，q_i，$i=1, 2, \cdots, m$，如$\sum\limits_{i=1}^{m} q_i=1$，有不等式

$$\sum_{i=1}^{m} q_i \cdot \log q_i \leqslant \sum_{i=1}^{m} q_i \cdot \log p_i \tag{2.9}$$

成立，其中等号成立的充要条件是$p_i=q_i$，$i=1, 2, \cdots, m$。

该引理的证明只要将(2.8)式中的q_i换为p_i/q_i就可，这时

$$\sum_{i=1}^{m} q_i \cdot \log(p_i/q_i) \leqslant \log\left[\sum_{i=1}^{m} p_i \cdot (p_i/q_i)\right]$$

$$=\log\left[\sum_{i=1}^{m} p_i\right]=0,$$

此即(2.9) 式成立。

命题1 对上述仙农熵$H(X)=H(p_1, \cdots, p_m)$总有

$$0\leqslant H(p_1, \cdots, p_m)\leqslant \log m$$

成立，其中第二个不等式成立的充要条件是$p_1=p_2=\cdots=p_m=1/m$， 而第一个不等式成立的充要条件是随机试验X是一个决定性试验，也就是有一个p_i为1，其余为零。

这个命题告诉我们，当随机试验X为均匀分布时不肯定性(仙农熵)为最大，它的证明由引理3即得，这时只要取(2.9)中的$p_i=1/m$，$i=1, \cdots, m$，就可。

由引理 1—3 的不等式还可得到关于信息量一系列其它性质，对此我们在以后的章、节中进一步讨论。

3.关于信息度量的研究状况

自仙农熵产生以后，信息的编码理论很快得到确立，因为由仙农熵很快可以派生出条件熵、交互信息、信号体积、信道容量等一系列概念。这些度量已成为信源、信道编码问题的理论基础，它们的定义与应用我们在以下各章中介绍。

如何扩展信息度量的范围一直是信息论工作者追求的目标，因为任何新的信息量的诞生就意味着新的信息科学分支的产生。例如，率失真函数的产生导致的数据压缩理论的建立，信道容量区域的产生导致多用户信息论的发展，对此我们就不一一列举了。

自仙农熵产生之后，在60—70年代，信息量的推广主要集中在从离散概率分布推广到连续型概率分布及从对数函数推广到一般下凸函数上。关于概率分布的推广苏联学派进行了大量的工作，他们采用了互熵的办法解决了由仙农熵直接推广的困难，对此我们在第四章中作详细的介绍。关于对数函数的推广

也有许多工作进行讨论，一般可采用幂函数等办法解决，一系列新的熵，如α-阶熵的产生就为此例。值得指出的是，这些新的信息量的产生虽大大丰富了信息论的内容，在信息统计等理论中也有许多应用，但它们的影响总不如仙农熵，其中的一些概念还可看作仙农熵的外延或派生。

与仙农熵出发点完全不同的两种度量是计算机科学中的描述(或程序)复杂性与分形几何中的豪斯道夫(Hausdorff)维数。描述复杂性又名柯尔莫各洛夫复杂性，以柯尔莫各洛夫复杂性为基础建立起来的信息量并由此建立的信息论为算法信息论。算法信息论的观点否定了以概率分布为基础的信息度量原则，认为不同事物有不同的信息度量，这个度量的大小应以计算的复杂性来统一。因此算法信息论是计算机科学与信息科学的结合点。关于豪斯道夫维数问题是一种自然界中自相似图形的结构度量，它反映了这种自相似图形的内在复杂性。因此在一定的意义下，柯尔莫各洛夫复杂性，豪斯道夫维数与仙农熵之间又存在一系列的等价关系，它们都可以说是客观世界中事物结构复杂性的表征。如果我们从这样的角度来理解信息的度量，那么我们就可以对信息、信息的度量与信息科学有更深入的理解。有关算法信息论与分形几何理论及他们与仙农熵的关系问题我们在第六章中再作介绍。

§1.3 信息论与信息科学

在以上的论述中我们已大体刻划了信息论的产生与发展的历史，以及与信息度量关系密切的学科分支。我们认为这些内容是近代信息论的组成部分，当然也是信息科学的组成部分。

从广义的意义上来讲，信息科学应有更广泛的内容，如信号、图象、语音的一系列信息处理问题或计算机的语言编码等问题等。它们无疑是信息科学的研究内容，这些内容比较偏重实际数据与各种具体内容的处理，因此与前者还有一定的区别。我们认为这两部分内容都是信息科学的重要组成部分，它们相互推动与发展。明确这一点对我们学习、了解信息科学的全貌是有帮助的。

第二章 仙农熵与无噪声信源编码问题

§ 2.1 通信系统概述

一个通信系统由信源、信道、编码与译码等要素构成，我们先对它们的含意与记号进行以下的描述与说明.

1.信源

信源即消息的来源，我们通常发送的电报、电话、信函、电视图象等等都可看作信源的消息. 由于通信系统的服务对象要求不同，不同的通信系统有不同的信源特征. 例如不同的国家有不同的文字记号，不同的应用范围(民用或专用系统)有不同的用字范围与频率，因此一个信源可通过以下内容进行描述.

(1) 信源消息字母表$\boldsymbol{X}$. 即信源消息可能使用的全部符号，如中文电报中的消息字母表为全体汉字、数字与标点符号等，在英文电报中的消息字母表是全体英文字母、数字与标点符号等. 在数学上，消息字母表为一个有限集合$\boldsymbol{X}$，它的元素

为信源字母或输入消息字母，我们用小写字母x、y等表示。

（2）一个确定的信源除了上述消息字母表确定之外，每个消息字母的使用概率往往是确定的，例如常用的英文字母的使用概率由表1确定。

表1　　常用英文字母使用概率表

字母	使用概率	字母	使用概率	字母	使用概率	字母	使用概率
A	0.0856	*H*	0.0528	*O*	0.0797	*V*	0.0092
B	0.0139	*I*	0.0627	*P*	0.0199	*W*	0.0149
C	0.0297	*J*	0.0013	*Q*	0.0012	*X*	0.0017
D	0.0378	*K*	0.0042	*R*	0.0677	*Y*	0.0199
E	0.1304	*L*	0.0339	*S*	0.0607	*Z*	0.0008
F	0.0289	*M*	0.0249	*T*	0.1045		
G	0.0199	*N*	0.0707	*U*	0.0249		

表2　　最常用的前十个汉字的概率表

编号	政治		文艺		新闻		科技		综合	
	字	概率	字	概率	字	概率	字	概率	字	概率
1	的	0.0536	的	0.0324	的	0.0375	的	0.0320	的	0.0384
2	是	0.0165	一	0.0218	一	0.0132	一	0.0097	一	0.0125
3	一	0.0136	了	0.0196	了	0.0120	在	0.0092	是	0.0098
4	在	0.0115	不	0.0165	和	0.0086	用	0.0079	在	0.0095
5	这	0.0109	是	0.0141	在	0.0086	有	0.0073	了	0.0082
6	主	0.0108	说	0.0130	人	0.0083	是	0.0070	不	0.0081
7	不	0.0101	他	0.0130	大	0.0083	不	0.0069	和	0.0075
8	和	0.0098	这	0.0119	主	0.0083	中	0.0066	有	0.0069
9	人	0.0087	着	0.0107	是	0.0078	大	0.0064	大	0.0069
10	们	0.0087	个	0.0097	们	0.0065	时	0.0063	这	0.0064
累计		0.1544		0.1627		0.1189		0.0922		0.1141

对汉字的统计工作是十分浩繁的，1981年我国公布了《信息交换用汉字编码字符基本集》的国家标准(GB-2312)，该标准收容字数为6763个，它为我国汉字信息处理的依据。表2、表3给出了汉字概率分布的特征表。

表3　　常用汉字分布情况表

累计概率	汉字序号				
	政治	文艺	新闻	科技	综合
0.50	102	96	132	169	163
0.90	650	860	780	900	950
0.99	1790	2180	2080	2250	2400
0.999	2996	3204	3402	3719	3804
0.9999	3917	3808	4575	5116	5265
1.0000	4356	3956	5084	5711	6359

对一般消息字母x的概率我们用$p(x)$来表示，这时记

$$\boldsymbol{S}=[\boldsymbol{X},p(x)] \tag{1.1}$$

为信源。如果X为一个在$\boldsymbol{X}$上取值且具有概率分布$p(x)$的随机变量，也就是$p_r\{X=x\}=p(x)$，那么我们称X为信源随机变量或输入消息随机变量。

2.信道

在通信过程中，信息通过信号荷载，信号经过的通道为信道传递。信道一般由甲、乙两地及一定的联系媒质组成，甲、乙两分别为输入、输出端或发送、接收端。在信道的输入端，电磁波以一定的脉冲信号经过一定的媒质进行发送，到达接收端，且变为输出信号进行接收。因此信道的组成内容有：

(1) 输入信号字母表$\boldsymbol{U}$。信道的全体可能的输入信号字母

(或符号)，在近代通信中常用q-进制的脉冲信号，因此$\boldsymbol{U}$常取为有限域(或整数集合)$\boldsymbol{GF}(q)=\{0,1,\cdots,q-1\}$.

(2) 输出信号字母表$\boldsymbol{V}$. 信道输出端的全体可能的输出信号字母(或符号)，一般取$\boldsymbol{U}=\boldsymbol{V}$，但也可以不相同.

如$\boldsymbol{U}$、$\boldsymbol{V}$都是有限集合，那么这个信道为离散信道，如$\boldsymbol{U}=\boldsymbol{V}=\boldsymbol{GF}(q)$，那么这个信道为一个$q$-进制信道. 在实用通信中，有时取$\boldsymbol{U}$为有限集合而$\boldsymbol{V}$为连续集合(如全体实数)，这时称这个信道为半连续信道.

(3) 信道噪声或干扰.当脉冲信号电磁波在信道中传递时，由于通信内部设备或信道外部条件的影响，原发送信号的脉冲形状会发生畸变，从而导致接收者的判断差错，人们统称这种畸变为信道的噪声或干扰，它可能的结果用一个条件概率分布$p(v/u)$来表示，它表示在发送信号u固定的条件下，接收信号为v的概率，这时

$$p(v/u)\geqslant 0,\ 且\ \sum_{v\in V}p(v/u)=1.$$

以下我们记

$$\boldsymbol{C}=[\boldsymbol{U},\ p(v/u),\ \boldsymbol{V}] \tag{1.2}$$

为信道. 如果$(U,\ V)$为两个随机变量且它们的条件概率为$p(v/u)$，也就是$P_r\{V=v/U=u\}=p(v/u)$，那么我们分别称U、V为信道的输入、输出信号随机变量.

3.编码

由信源消息变为信道发送信号的过程为编码. 一个实用通信系统的编码过程是由一系列符号之间的运算构成，如在汉字的电报通信中，编码过程为

汉字⇒数字⇒脉冲信号⇒带电磁波的脉冲信号.

信道实际上发送的是带电磁波的脉冲信号，在工程中把以上变换过程称为调制过程．在信息论中常常把“汉字⇒脉冲信号”的运算过程称为编码．在一般情形，编码就是一个从信源字母表$\boldsymbol{X}$到信道输入信号字母表$\boldsymbol{U}$的变换．我们记之为f_E．

4.译码

由信道接收信号变为信源消息的过程为译码．因此译码过程是编码的逆过程，实用通信系统的译码过程是一个解调过程，如以上汉字电报通信中的解调过程是

带电磁波的脉冲信号⇒脉冲信号⇒数字⇒汉字．

由于信道的干扰影响，接收信号会有一定程度的畸变发生，因此解调过程首先是一个从带电磁波的接收信号到脉冲接收信号的一个统计判决过程，这个判决结果产生信道转移概率$p(v/u)$．在信息论中的译码就是指由脉冲信号到汉字的变换运算．在一般情形就是一个从信道输出信号字母表$\boldsymbol{V}$到信源复制字母表$\boldsymbol{Y}$的变换运算，我们记之为f_D．信源复制字母表$\boldsymbol{Y}$即$\boldsymbol{X}$的复制空间，一般取$\boldsymbol{Y}=\boldsymbol{X}$，有时也可不同．

在信息论的术语中常把以上定义的“编码”与“译码”合称编码，有时记(f_E,f_D)为(f,g)．

5.通信系统

综上所述，一个通信系统是由信源、信道、编码与译码等要素构成．为了以后叙述的方便，对不同要素的组成的系统给以不同名称．以下我们记

$$\boldsymbol{E}=\{\boldsymbol{S},\boldsymbol{C}\}=\{[\boldsymbol{X},p(x)],[\boldsymbol{U},p(v/u),\boldsymbol{V}]\} \quad (1.3)$$

为通信系统．这是通信工程中最为固定的成分，因为一般的通信环境与硬件设备是不会轻易改变的．这时，信源的概率分布

与信道的条件概率分布都是由客观条件所确定．而称

$$\boldsymbol{E}(f_E,f_D)=\{\boldsymbol{E},(f_E,f_D)\}=\{\boldsymbol{S},\boldsymbol{C},(f_E,f_D)\} \quad (1.4)$$

为具有编码的通信系统．因为编码与译码函数可以人为选择，它们的调整的技术上也比较容易实现，因此仙农信息论的基本特点就是通过研究编码与译码的构造来实现最佳通信．

把一个同时具有信源、信道、编码、译码、调制与解调的系统称为通信工程系统．因为调制与解调问题涉及其它一系列工程与数学问题，对此本书不作详述．

在信息论中，一个通信系统可用框图2.1来表示．如通信系统$\boldsymbol{E}(f_E,f_D)$给定，那么$\boldsymbol{X}\times\boldsymbol{U}\times\boldsymbol{V}\times\boldsymbol{Y}$上的联合概率分布由

$$p(x,u,v,y)=p(x)\cdot f_E(u/x)\cdot p(v/u)\cdot f_D(y/v) \quad (1.5)$$

完全确定，其中

$$f_E(u/x)=\begin{cases}1, & \text{如果} f_E(x)=u,\\ 0, & \text{否则},\end{cases}$$

$$f_D(y/v)=\begin{cases}1, & f_D(v)=y,\\ 0, & \text{否则}.\end{cases} \quad (1.6)$$

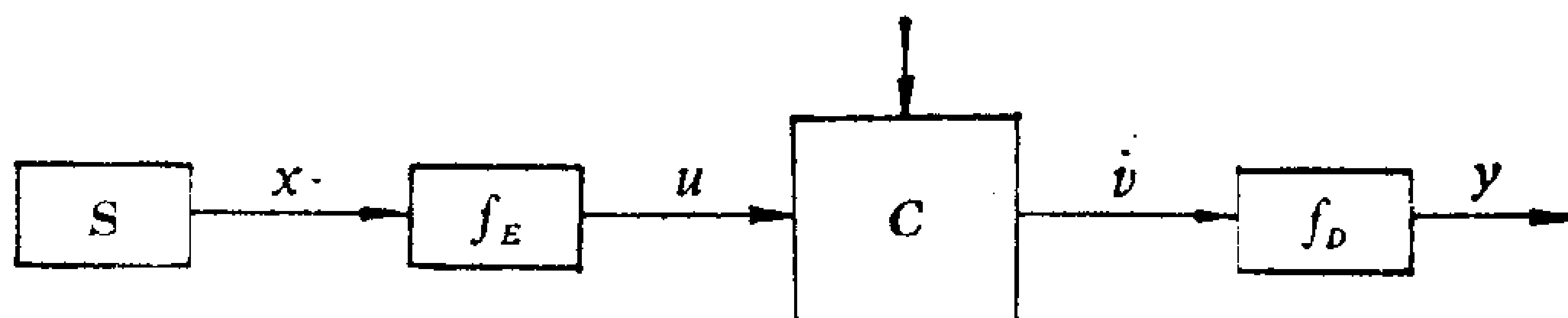


图2.1 通信系统的框图

如果随机变量(X, U, V, Y)取值于$\boldsymbol{X}\times\boldsymbol{U}\times\boldsymbol{V}\times\boldsymbol{Y}$，而且它的概率分布为由(1.5)式给定，也就是，对任何$(x,u,v,y)\in\boldsymbol{X}\times\boldsymbol{U}\times\boldsymbol{V}\times\boldsymbol{Y}$，有

$$P_r\{X=x, U=u, V=v, Y=y\}=p(x, u, v, y) \quad (1.7)$$

成立，其中$p(x,u,v,y)$由(1.5)定义，这时我们称(X,U,V,Y)为由通信系统$\boldsymbol{E}(f_E, f_D)$决定的随机变量，有时直接称$\{X,U,V,Y\}$为通信系统。当$X=Y$时，我们称：

$$p_e(f_E,f_D)=P_r\{X\neq Y\}=\sum_{(x,y):x\neq y} p(x,y) \quad (1.8)$$

为通信系统的概率误差，其中$p(x,y)$为$p(x,u,v,y)$在$\boldsymbol{X}\times\boldsymbol{Y}$上的边际分布，也就是

$$p(x,y)=P_r\{X=x,Y=y\}=\sum_{v\in V}\sum_{u\in U} p(x,u,v,y). \quad (1.9)$$

6.通信系统的序列模型

在实际的通信问题中，无论是消息字母或信号字母都是成串传递的。因此实际的传递消息、信号是一组向量，它们分别是

输入消息向量：$x^n=(x_1,x_2,\cdots,x_n)$，$x_i\in\boldsymbol{X}$，$i=1,2,\cdots,n$，

输入信号向量：$u^n=(u_1,u_2,\cdots,u_n)$，$u_i\in\boldsymbol{U}$，$i=1,2,\cdots,n$，

输出信号向量：$v^n=(v_1,v_2,\cdots,v_n)$，$v_i\in\boldsymbol{V}$，$i=1,2,\cdots,n$，

复制消息向量：$y^n=(y_1,y_2,\cdots,y_n)$，$y_i\in\boldsymbol{Y}$，$i=1,2,\cdots,n$，

它们的向量空间分别是$\boldsymbol{X}^n$，$\boldsymbol{U}^n$，$\boldsymbol{V}^n$与$\boldsymbol{Y}^n$。相应的信源、信道分别是

$$\boldsymbol{S}^n=[\boldsymbol{X}^n,\ p(x^n)],\quad \boldsymbol{C}^n=[\boldsymbol{U}^n,\ p(v^n/u^n),\ \boldsymbol{V}^n], \quad (1.10)$$

其中$p(x^n)$是$\boldsymbol{X}^n$上的概率分布，而$p(v^n/u^n)$是条件概率分布．(1.10)中的上标n是向量长度，如$n=1,2,3,\cdots$，可以任意延长，那么我们称(1.10)的$\boldsymbol{S}^n$，$\boldsymbol{C}^n$分别为信源序列与信道序列，而称

$$\boldsymbol{E}^n=\{\boldsymbol{S}^n,\ \boldsymbol{C}^n\},\ \boldsymbol{E}^n(f_E^n,f_D^n)=\{\boldsymbol{E}^n,\ (f_E^n,f_D^n)\},$$
$$n=1,2,3,\cdots, \tag{1.11}$$

为通信系统序列，其中f_E^n，f_D^n分别为编码与译码运算，它们分别是

$$f_E^n:\ \boldsymbol{X}^n\to\boldsymbol{U}^n,\ f_D^n:\ \boldsymbol{V}^n\to\boldsymbol{Y}^n$$

的变换运算．仿照(1.6)—(1.9)的定义，同样可以确定通信系统序列的联合概率分布$p(x^n,\ u^n,\ v^n,\ y^n)$，通信系统的随机变量$\{X^n,\ U^n,\ V^n,\ Y^n\}$，及相应的误差概率$p_e(f_E^n,\ f_D^n)$，对此我们就不一一列举．

我们称信源序列$\boldsymbol{S}^n$，$n=1,2,\cdots$，是一个“无记忆的信源序列”，如果对任何$x^n=(x_1,\ \cdots,\ x_n)\in\boldsymbol{X}^n$，$n=1,2,3,\ \cdots$，总有

$$p(x^n)=\prod_{i=1}^{n}p(x_i) \tag{1.12}$$

成立．

同样我们称信道序列$\boldsymbol{C}^n$，$n=1,2,\cdots$，是一个“无记忆的信道序列”，如果对任何$u^n=(u_1,\cdots,u_n)\in\boldsymbol{U}^n$，$v^n=(v_1,\cdots,v_n)\in\boldsymbol{V}^n$，$n=1,2,3,\cdots$，总有

$$p(v^n/u^n)=\prod_{i=1}^{n}p(v_i/u_i) \tag{1.13}$$

成立．

无记忆信源、信道是通信系统中最常见的通信模型，它不仅具有广泛的实际背景，且对它们的编码问题的讨论包含了信息论的基本特征．对有记忆情形，我们只要利用随机过程的工

具作相应的推广就可。因此，在本书中，我们讨论的通信模型以无记忆信源、信道为基础。

仙农信息论的基本问题就是在信道有干扰或无干扰的条件下实现最佳通信的要求。所谓最佳通信要求就是在实现正确无误差的通信前提下，尽可能多地传送信源信息。依据信源、信道的结构特点，人们常把以上基本问题分为信源、信道编码问题，这些问题的解决与仙农熵密切相关。

§2.2 信源的不等长信号编码问题

无噪声信源编码问题是一种最常见的编码问题，它的一般定义是指把一种字母表的字母变为另一种字母表的字母。例如计算机中的8421码，它就是把一个十进位数变成一个 4 位二进位数。这种编码是在信道无干扰条件下使用。

无噪声信源编码的类型很多，一般分为等长与不等(或变)长编码及有允许误差与不允许误差等类型。在本章中，我们只讨论不允许误差情形。对有允许误差的编码问题在第五章中有详细讨论。在本节中，我们讨论信源的无噪声变长码问题。

1.无噪声信源不等长编码的一般定义

如§2.1所记，设 $\boldsymbol{X}$ 为信源消息字母表，$\boldsymbol{U}$ 为信源信号字母表，而

$$\boldsymbol{X}^n=\prod_{i=1}^{n}\boldsymbol{X}_i,\quad \boldsymbol{U}^n=\prod_{i=1}^{n}\boldsymbol{U}_i, \tag{2.1}$$

分别为它们的n-维乘积空间，其中$\boldsymbol{X}_i=\boldsymbol{X}$，$\boldsymbol{U}_i=\boldsymbol{U}$. 以下记

$$\boldsymbol{X}^*=\bigcup_{n=1}^{\infty}\boldsymbol{X}^n,\quad \boldsymbol{U}^*=\bigcup_{n=1}^{\infty}\boldsymbol{U}^n, \tag{2.1'}$$

这时，$\boldsymbol{X}^*$、$\boldsymbol{U}^*$分别为$\boldsymbol{X}$、$\boldsymbol{U}$上的全体不等长向量．信源的不等长编码是指一个从$\boldsymbol{X}$到$\boldsymbol{U}^*$的变换．我们称

$$\boldsymbol{U}_f=\{f(x),\ x\in\boldsymbol{X}\}\subset\boldsymbol{U}^* \tag{2.2}$$

为一个变长码．

定义1 对变长码$\boldsymbol{U}_f$，我们定义：

(1) $f(x)$是一个1－1码，如果$x\neq x'$，那么$f(x)\neq f(x')$。

(2) $f(x)$是一个唯一可译码，如果对任何x^n，$x'^{n'}\in\boldsymbol{X}^*$，如果$x^n\neq x'^{n'}$，那么必有$f(x^n)\neq f(x'^{n'})$，其中

$$x^n=(x_1,\ x_2,\ \cdots,\ x_n),\quad x'^{n'}=(x_1',x_2',\cdots,x'_{n'}),$$

其中$x^n\neq x'^{n'}$是指$n\neq n'$或$x^n\neq x'^n$，而

$$f(x^n)=(f(x_1),\ f(x_2),\ \cdots,\ f(x_n)). \tag{2.3}$$

(3) $f(x)$是一个前缀(prefix)码，如果对任何$x,x'\in\boldsymbol{X}$，$f(x)$与$f(x')$相互不为前缀。

对以上定义我们有以下几点说明：

(1) 关于前缀向量可由以下例子说明，如讨论向量：

$$a=(0,1,0),\ b=(0,1,0,0,0,1),\ c=(1,0),$$

这时向量a是b的前缀，而c既不是a也不是b的前缀。

(2) 关于1－1码与唯一可译码的关系是：唯一可译码一定是1－1码，反之则不然．此关系的正命题是显然的，因为满足唯一可译码的条件必满足1－1码的条件．对它的反命题不成立可由以下例子说明。

如我们取$\boldsymbol{X}=\{1,2,3\}$，$\boldsymbol{U}=\{0,1\}$，而取

$$f(1)=0,\ f(2)=1,\ f(3)=(0,1).$$

这时$f(x)$是1－1码，因为如$x\neq x'$时，必有$f(x)\neq f(x')$．但$f(x)$不是唯一可译码，因为$(1,2)\neq 3$，但是有

$$f(1,2)=(f(1),\ f(2))=(0,1)=f(3).$$

这时$\boldsymbol{U}_f$不是唯一可译码。

(3) 关于前缀码与唯一可译码的关系由以下定理说明。

定理1 任何前缀码一定是唯一可译码。

证明：如果$f(x)$是任一前缀码，那么它一定是一个唯一可译码；不然，总有一对$x^n \neq x'^{n'} \in \boldsymbol{X}^*$，使得

$$\begin{aligned} f(x^n)&=(f(x_1),\ f(x_2),\ \cdots,\ f(x_n)) \\ &=f(x'^{n'})=(f(x_1'),\ f(x_2'),\ \cdots,\ f(x'_{n'})) \end{aligned} \tag{2.4}$$

成立。因为$f(x)$为前缀码，这时必有$x_1=x_1'$成立，否则在(2.4)中，必有$f(x_1)$为$f(x_1')$(或$f(x_1')$为$f(x_1)$)的前缀。这时(2.4)式变为有

$$\begin{aligned} &(f(x_2),\ f(x_3),\ \cdots,\ f(x_n)) \\ &=(f(x_2'),\ f(x_3'),\cdots,\ f(x'_{n'})) \end{aligned} \tag{2.4'}$$

成立。由(2.4′)同样可得$x_2=x_2'$成立，依次类推，可得$x^n=x'^{n'}$成立。这与假定矛盾，因此$f(x)$一定是唯一可译码。定理得证。

(4) $\boldsymbol{U}^n$的树结构

我们现在讨论向量空间$\boldsymbol{U}^n$的一种树图表示。不失一般性，我们取$\boldsymbol{U}=\{0,1\}$。我们对照图2.2对树图进行描述定义。由图2.2可知，树图由若干节点与线段组成，由每个节向右产生树的分杈，我们称最起始的节点为根点。连结节点的线段为树的节，不同的节用$\boldsymbol{U}$中不同的值来表示，我们称之为节的状态，由同一个节点分杈的节的状态互不相同。

在树图中，由根点出发沿不同分杈向右可得到不同的枝，每条枝中所含节的数目为枝的长。图2.2中长度为4的枝有2^4条，每条长为4的枝正是$\boldsymbol{U}^4$中不同的向量。这时我们称图2.2为(2,4)

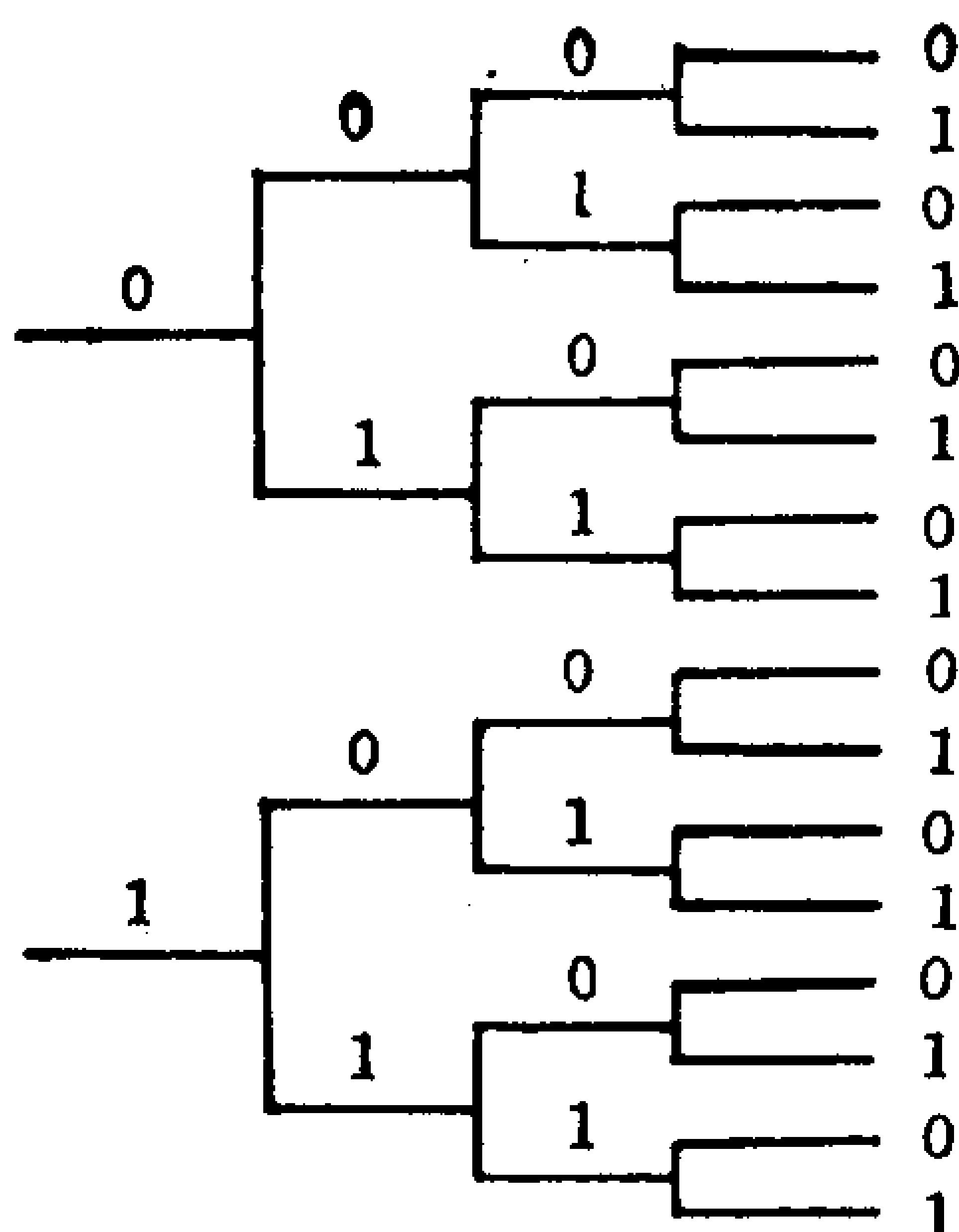


图2.2 (2,4)—完全树

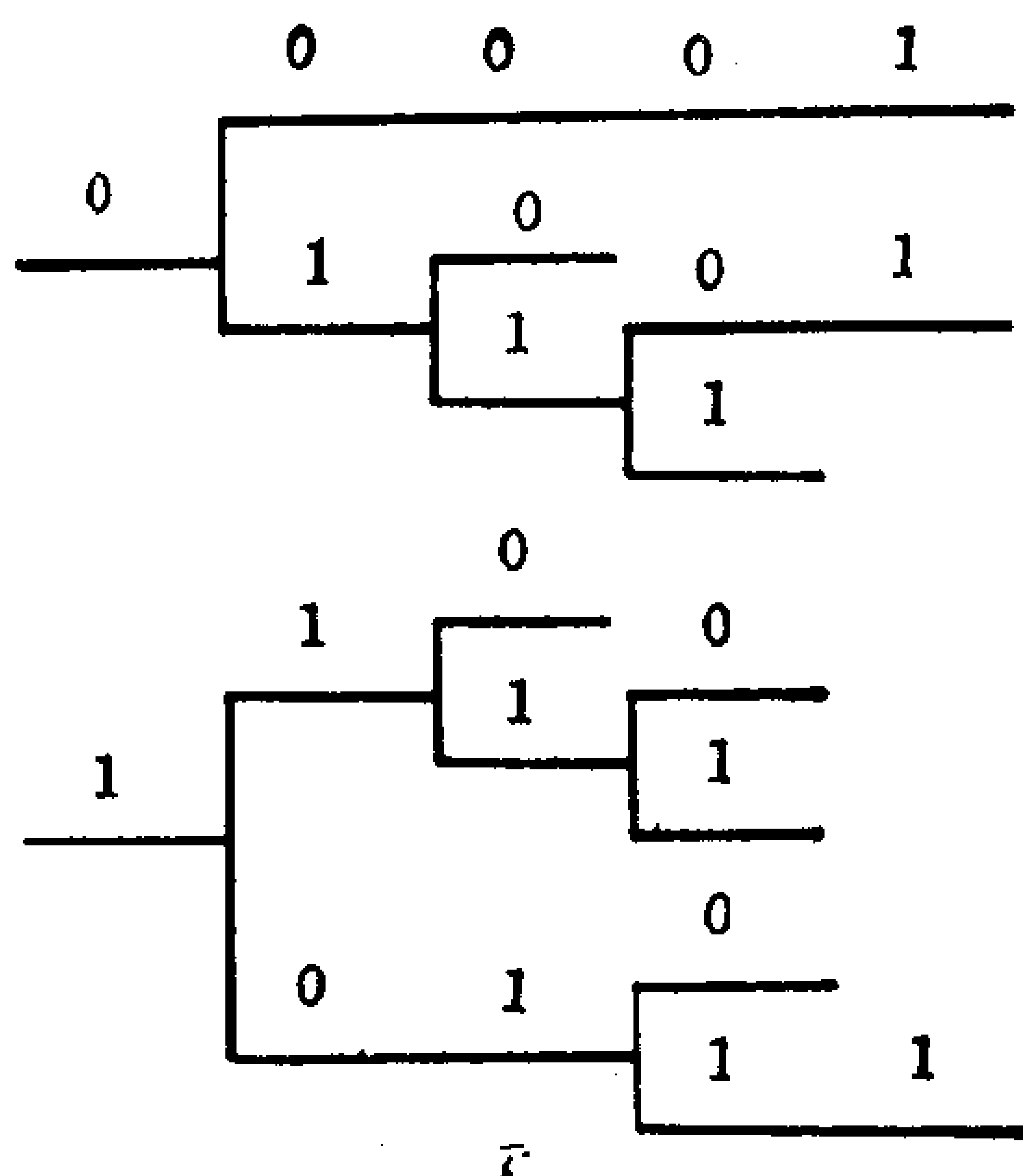


图2.3 $\boldsymbol{U}_f$不完全树

-完全树，因为它的全部长度为4的枝正是$\boldsymbol{U}^4$码的全部向量。

在一般情形，一个$\boldsymbol{U}^n$空间为一个(D, n)-完全树，每个节点有D个分权，其中$D=\|\boldsymbol{U}\|$为$\boldsymbol{U}$的元素个数。因此，在(D,n)-完全树中，长度为n的枝有D^n条。

一个不等长码$\boldsymbol{U}_f$为一个(D,n)-完全树的子树，它由一个(D,n)-完全树的若干枝上剪去部分分权而得，其中n为$\boldsymbol{U}_f$中最长的向量长度。例如，当

$$\boldsymbol{U}_f=\{(010),(01101),(0111),(00001),(10111),\\(1110),(110),(1010),(1111)\}$$

时，它的树结构图如图2.3所示。

由前缀码的定义可知，$\boldsymbol{U}_f$为前缀码的充要条件是在$\boldsymbol{U}_f$的树结构表示中，由根点到达树的末梢点的枝有$\|\boldsymbol{X}\|$条。例如，在图2.3中，到达树的末梢点的枝数与码元数都为9，因此$\boldsymbol{U}_f$为前缀码。

(5) 前缀码的构造条件与方法可由以下定理给定。

定理2 （L. G. Kraft不等式）。如我们记 $X=\{x_1, x_2, \cdots, x_k\}$，$D=\|\boldsymbol{U}\|$，$n_1$，$n_2$，…，$n_k$为一组正整数．那么有：如果不等式

$$\sum_{j=1}^{k} \exp_D(-n_j) \leqslant 1, \tag{2.5}$$

成立，那么必可构造一个前缀码$f(x)$，使

$$n_j = l[f(x_j)], \quad j=1,2,\cdots,k, \tag{2.6}$$

成立，其中$l[f(x)]$为变长码$f(x)$的码长．反之，如果$f(x)$为前缀码，那么它的码长n_j，$j=1,2,\cdots,k$必满足(2.5)式．

证明：我们先证逆命题．如果$\boldsymbol{U}_0=\{f(x), x\in\boldsymbol{X}\}$是一个前缀码，那么(2.5) 式必成立．如我们记$N=\max\{n_i, j=1,2,\cdots,k\}$，且记

$$u^{t+s}=(u_1,\cdots,u_t,u_{t+1},\cdots,u_{t+s})=(u_1^t,u_2^s)$$

为两个向量的联结，其中

$$u_1^t=(u_1, \cdots, u_t), \quad u_2^s=(u_{t+1}, \cdots, u_{t+s}).$$

这时记$u^t\odot\boldsymbol{U}^s=\{(u^t, u^s): u^s\in\boldsymbol{U}^s\}$为由枝$u^t$产生的全部长度为$t+s$的枝。这时 $\|u^t\odot\boldsymbol{U}^s\|=\exp_D(s)$。如果$f(x)$是一个前缀码，那么

$$f(x_i)\odot\boldsymbol{U}^{N-n_i}, \quad i=1,2,\cdots,k, \tag{2.7}$$

是一组长度为N的，互不相同的向量族，向量的总数为

$$\sum_{i=1}^{k} \exp_D(N-n_i) \leqslant \exp_D(N). \tag{2.8}$$

把(2.8)式两边同除以$\exp_D(N)$即得(2.5)式成立，逆命题得证。

如果(2.5)式成立，我们可构造满足(2.6)式的前缀码．不失一般性，我们取$n_1\leqslant n_2\leqslant\cdots\leqslant n_k=N$，因为$\boldsymbol{U}^N$为一个$(D,N)$-完全树．这时，前缀码$\boldsymbol{U}_f$可由完全树$\boldsymbol{U}^N$剪接而成，剪接步骤如下：

(1) 从根点出发，任取一长度为n_1的枝，我们记它的状态为

$$u_1^{n_1}=(u_{1,1},\ u_{1,2},\ \cdots,\ u_{1,n_1})。$$

剪去由$u_1^{n_1}$枝所产生的全体分杈。剪去的分枝有$\exp_D(N-n_1)$条，这时完全树$\boldsymbol{U}^N$还余下$\exp_D(N)-\exp_D(N-n_1)$条长度为N的枝，我们记之为$\boldsymbol{U}_1$。

(2) 从根点出发，从$\boldsymbol{U}_1$中任取一长度为n_2的枝，我们记之为

$$u_2^{n_2}=(u_{2,1},u_{2,2},\cdots,u_{2,n_2})。$$

同样剪去由$u_2^{n_2}$枝所产生的全体分杈。剪去的分枝有$\exp_D(N-n_2)$条，这时完全树$\boldsymbol{U}^N$还余下$\exp_D(N)-\sum\limits_{i=1}^{2}\exp_D(N-n_i)$条长度为$N$的枝，我们记之为$\boldsymbol{U}_2$。

(3) 依次类推，我们可以找到一系列的枝

$$u_i^{n_i}=(u_{i,1},u_{i,2},\cdots,u_{i,n_i}),\ i=1,2,3,\cdots, \qquad (2.9)$$

因为有(2.5)成立，(2.9)中的$u_i^{n_i}$至少有k条。这时我们取

$$f(x_i)=u_i^{n_i},\ i=1,2,\cdots,k,$$

就为所求的前缀码，且$l[f(x_i)]=n_i$，$i=1,2,\cdots,k$，成立。定理得证。

2.信源的不等长信号编码定理

所谓信源的不等长信号编码问题就是讨论对已给信源$X=[\boldsymbol{X},p(x)]$，如何构造平均长度较短的唯一可译码。如上所记，设$f(x)$是$\boldsymbol{X}$上的一个不等长编码，这时$f(x)$是一个$\boldsymbol{U}^*$中的向量，它的长度记为$l[f(x)]$，而记

$$L[f(X)]=\boldsymbol{E}\{l[f(X)]\}=\sum_{x\in X}l[f(x)]\cdot p(x) \qquad (2.10)$$

为变长码$f(x)$的平均长度，其中，$\boldsymbol{E}\{Y\}$ 表示随机变量Y的数学期望(均值)。以下我们讨论$f(x)$的平均长度。

定理3 任何唯一可译码$f(x)$，必有

$$L[f(X)]\geqslant H(X)/(\log D), \tag{2.11}$$

其中 $H(X)$为随机变量X 的熵。反之，必存在适当的唯一可译码$f(x)$，使

$$L[f(X)]\leqslant H(X)/(\log D)+1 \tag{2.12}$$

成立。

证明：我们先证(2.11)式。在定理2的记号下，又记

$$q_i=\exp_D(n_i),\ \ p_i=p(x_i),\ \ i=1,2,\cdots,k.$$

这时有$-\log(q_i)=n_i\cdot\log D$，且有

$$H(X)-L[f(X)]\cdot\log D=\sum_{i=1}^{k}p_i\cdot\log(q_i/p_i)$$

$$\leqslant\sum_{i=1}^{k}p_i\cdot(q_i/p_i-1)\cdot\log_e(2)$$

$$=\sum_{i=1}^{k}(q_i-p_i)\cdot\log_e(2)\leqslant 0. \tag{2.13}$$

在(2.13)中，第一个不等式由对数函数性质：$\log x\leqslant(x-1)\cdot\log_e(2)$而得，而第二个不等式由定理2即得.(2.11)式得证。

对(2.12)式，我们取

$$-[\log p_i/\log D]\leqslant n_i<-[\log p_i/\log D]+1, \tag{2.14}$$

这时n_i，$i=1,2,\cdots,k$,必满足定理2的 (2.5)式条件，因此由定理2可构造一个前缀码$f(x),x\in\boldsymbol{X}$，使$n_i=l[f(x_i)]$。这时有

$$L[f(X)]=\sum_{i=1}^{k}n_i\cdot p_i$$

$$<\sum_{i=1}^{k}[-(\log p_i/\log D)+1]\cdot p_i=H(X)/\log D+1,$$

此即(2.12)式成立。定理得证。

由定理2，3给出了唯一可译码(或前缀码)的构造方法与平均长度的变化范围。这些性质反映了信源的无噪声、不等长编码的基本性质，为通信理论的基本特征之一。

§2.3 熵功率与信号体积

在§2.2中，我们给出了信源的不等长信号编码问题，它的唯一可译码的平均长度为该信源的熵控制。我们现在讨论信源的等长编码问题，它的特征可形象地用"熵功率"与"信号体积"来说明。

一个信源的等长编码问题是指编码函数为

$$f^n(x^n):\ \boldsymbol{X}^n \to \boldsymbol{U}^m \tag{3.1}$$

的映射．当 $(n,\ m)$ 固定时，如果 $f^n(x^n)$ 是个1－1映射，那么 $f^n(x^n)$ 一定是唯一可译的．但是，在信源的等长编码问题中，因为当上标n较大时，$\boldsymbol{X}^n$是一个十分巨大的空间，许多向量在通信时实际上不会使用．因此，在信源的等长编码问题中，我们只要求 $f^n(x^n)$ 能以很大的概率还原就可．不失一般性，在本节中我们取$\boldsymbol{U}^n=\{0,1\}$，$m=R\cdot n$，这时称R为编码函数$f^n(x^n)$的码率．在本节中，我们讨论的信源模型是§2.1中的无记忆信源序列模型．它的概率分布由(1.12)式给定．

定义1 (1) 我们称R是信源$\boldsymbol{S}$的一个"可达率"，如果对任何$\varepsilon>0$，只要n充分大，就有一对编码函数 $(f^n,\ g^n)$，它们分别是

$$f^n(x^n):\ \boldsymbol{X}^n \to \boldsymbol{U}^m,\quad g^n(u^m):\ \boldsymbol{U}^m \to \boldsymbol{X}^n \tag{3.2}$$

的变换，其中$m=R\cdot n$，且编码的误差概率

$$p_e(f^n, g^n) = P_r\{g^n[f^n(X^n)] \neq X^n\} < \varepsilon. \qquad (3.3)$$

(2) 我们称R是信源$\boldsymbol{S}$的一个“最小可达率”，如果R是$\boldsymbol{S}$的可达率，且对任何$R' < R$，R'不为信源$\boldsymbol{S}$的可达率。

定理1 无记忆信源$\boldsymbol{S}$的最小可达率是$R_0 = H(X)$。

证明：对此定理我们分以下几步证明。

(1) 先证定理的正命题。也就是，当$R > R_0$时，R为信源的可达率。如我们定义

$$h(x^n) = \log p(x^n) = \sum_{i=1}^{n} \log p(x_i) \qquad (3.4)$$

为无记忆信源的“熵密度”，其中$p(x^n)$由(1.12)定义。这时，我们可把$h(x^n)$看作一个独立随机变数之和，由大数定律可知

$$P\text{-}\lim_{n\to\infty}(1/n)\cdot h(x^n) = \boldsymbol{E}\{h(X)\} = H(X) \qquad (3.5)$$

成立，其中“P-lim”为以概率收敛，$h(x) = -\log p(x)$为$p(x)$的熵密度。这时，对任何$\varepsilon > 0$，我们定义

$$\boldsymbol{X}_0^n = \{x^n: \ |(1/n)\cdot h(x^n) - H(X)| < \varepsilon\}. \qquad (3.6)$$

这时对任何$x^n \in X_0^n$，必有

$$\exp\{-n\cdot[H(X)+\varepsilon]\} < p(x^n) < \exp\{-n\cdot[H(X)-\varepsilon]\}$$

成立。因此有

$$\|\boldsymbol{X}_0^n\| < \exp\{n\cdot[H(X)+\varepsilon]\} < \exp[n\cdot(R+\varepsilon)] \qquad (3.7)$$

成立。由(3.5)式可知，对上述$\varepsilon > 0$，只要n充分大，就有

$$p(\boldsymbol{X}_0^n) = \sum_{x^n \in X_0^n} p(x^n) > 1 - \varepsilon \qquad (3.8)$$

成立。这样我们可在$\boldsymbol{X}_0^n$与$\boldsymbol{U}^m$之间构造一个1－1对应的函数$Q(x^n)$，其中$m = n\cdot(R+\varepsilon)$，且构造编码函数为

$$f^n(x^n) = \begin{cases} Q(x^n), & \text{如} x^n \in \boldsymbol{X}_0^n, \\ \text{任取一向量}, & \text{如} x^n \in \boldsymbol{X}_0^n, \end{cases} \qquad (3.9)$$

$$g^n(u^n)=\begin{cases}Q^{-1}(u^m), & 如u^m\in \boldsymbol{U}_0^m,\\ 任取一向量, & 如u^m\in \boldsymbol{U}_0^m,\end{cases} \tag{3.9'}$$

其中$Q^{-1}(u^m)$为$Q(x^n)$的逆函数，而

$$\boldsymbol{U}_0^m=\{u^m=Q(x^n):\ x^n\in \boldsymbol{X}_0^n\}.$$

这时

$$p_e(f^n,\ g^n)\leqslant 1-p(\boldsymbol{X}_0^n)<\varepsilon \tag{3.10}$$

成立。定理1的正命题得证。反之，如果$R<R_0$，那么我们取$\varepsilon_0=(R_0-R)/4$，对(3.7)的$\boldsymbol{X}_0^n$，取$\varepsilon=\varepsilon_0$，当$n$充分大时(3.8)式成立。如记

$$\boldsymbol{X}_1^n=\boldsymbol{X}^n-\boldsymbol{X}_0^n.$$

这时$p(\boldsymbol{X}_1^n)<\varepsilon_0$。我们现在考虑编码函数

$$f^n:\ \boldsymbol{X}^n\to\boldsymbol{U}^m,\quad g^n:\ \boldsymbol{U}^m\to\boldsymbol{X}^n,$$

如果$m=(R+\varepsilon_0)\cdot n$，那么总有

$$p_e(f^n,g^n)>1/2 \tag{3.11}$$

成立。因为这时我们记

$$\boldsymbol{X}_2^n=\{x^n=g^n(u^m):\ u^m\in\boldsymbol{U}^m\},$$

那么$\|\boldsymbol{X}_2^n\|\leqslant\exp[n\cdot(R+\varepsilon_0)]$，且有

$$1-p_e(f^n,\ g^n)=P_e\{g^n[f^n(X^n)=X^n\}\leqslant p(\boldsymbol{X}_2^n)$$
$$\leqslant p(\boldsymbol{X}_1^n)+p(\boldsymbol{X}_0^n\cap\boldsymbol{X}_2^n)\leqslant\varepsilon_0+\exp(-n\cdot\varepsilon_0). \tag{3.12}$$

其中(3.14)的第二个不等式由

$$p(\boldsymbol{X}_0^n\cap\boldsymbol{X}_2^n)\leqslant\exp\{n\cdot[R+2\cdot\varepsilon_0-H(X)]\}$$
$$\leqslant\exp(-2\cdot n\cdot\varepsilon_0)$$

而得。因为$\varepsilon_0\leqslant 1/4$，且当$n$充分大时，可使$\exp(-2\cdot n\cdot\varepsilon_0)<1/4$成立，代入(3.14)式就得(3.13)式成立。定理得证。

由以上编码定理可知，对无记忆信源序列的消息以概率1集中在一个具有$v(n)=\exp[n\cdot H(X)]$个点的子集上。仙农形像

地把它称为“信号体积”，也就是，$v(n)$ 的大小是一个n-维空间中边长为 $\exp[H(X)]$ 的正方体的体积。这时$H(X)$ 被称为“熵功率”。

第三章　信道容量与信道编码定理

§3.1　条件熵与交互信息

1.多元随机变量的仙农熵

在§1.2中，我们已引进仙农熵$H(X)$的定义，给出了它的来源与性质。它作为一般随机变量（或概率场）X的不肯定性的度量被引进。因此，它可直接推广到多元随机变量的情形。

设$(X_1, X_2, \cdots, X_n)$为一个n-元随机向量，每个X_i取值于$\mathbf{X}_i$空间，且具有联合概率分布为

$$p(x_1, \cdots, x_n) = p_r\{(X_1, \cdots, X_n) = (x_1, \cdots, x_n)\}. \tag{1.1}$$

它的仙农熵为

$$H(X_1, X_2, \cdots, X_n)$$
$$= -\sum_{X_1} \cdots \sum_{X_n} p(x_1, \cdots, x_n) \log p(x_1, \cdots, x_n) \tag{1.2}$$

我们称之为n元的联合仙农熵或简称联合熵，其中“$\sum\limits_{X_i}$”为对全体$x_i \in \mathbf{X}_i$求和，对此在下文中还要使用，但我们不再说明。

特殊情形，对二元随机变量（X，Y），如它的联合概率分布为

$$p(x,\ y)=P_r\{(X,\ Y)=(x,\ y)\}, \tag{1.3}$$

那么它的联合熵为

$$H(X,\ Y)=-\sum_{x,y}p(x,\ y)\log p(x,\ y). \tag{1.4}$$

以下我们总是记$\mathbf{X}$、$\mathbf{Y}$为X、Y的值空间，且称$H(X,Y)$为二元熵。

引理1 对二元熵$H(X,Y)$与仙农熵$H(X)$，$H(Y)$有以下不等式

$$\max\{H(X),\ H(Y)\}\leqslant H(X,\ Y)\leqslant H(X)+H(Y) \tag{1.5}$$

成立，其中第一个不等式中等号成立的充要条件是X由Y，或Y由X确定(也就是，$Y=f(X)$或$X=f(Y)$以概率1 成立)。第二个不等式中等号成立的充要条件是 X 与 Y 是 相 互独立的随机变量。

证明：利用 § 1.2的不等式（2.3）可得

$$\begin{aligned}H(X,Y)&=-\sum_{x}\Big[\sum_{y}p(x,\ y)\log p(x,y)\Big]\\&\geqslant-\sum_{x}\Big\{\Big[\sum_{y}p(x,y)\Big]\log\Big[\sum_{y}p(x,y)\Big]\Big\}\\&=-\sum_{x}p(x)\log p(x)=H(X),\end{aligned} \tag{1.6}$$

且(1.6) 的等号成立的充要条件是对每一个$x\in\mathbf{X}$，$p(x)>0$，总有一个$y(x)\in\mathbf{Y}$，使

$$p(x,y)=\begin{cases}p(x), & \text{如果}y=y(x)\\0, & \text{否则}.\end{cases}$$

这就是随机变量Y由X确定。同理可得$H(X,Y)\geqslant H(Y)$，且等

号成立的充要条件是X由Y确定．由§1.2的(2.9)式可得

$$H(X,Y)\leqslant-\sum_{x}\sum_{y}p(x,y)\log[p(x)q(y)]$$

$$=-\sum_{x}p(x)\log p(x)-\sum_{y}q(y)\log q(y)$$

$$=H(X)+H(Y), \tag{1.7}$$

其中$p(x)=\sum_{y}p(x,y)$，$q(y)=\sum_{x}p(x,y)$，且等号成立的充要条件是对任何$(x,y)\in\boldsymbol{X}\times\boldsymbol{Y}$，有$p(x,y)=p(x)q(y)$成立。定理得证。

2.条件熵与交互信息

利用多元随机变量的联合熵即可引进条件熵与交互信息。

定义1 设(X，Y)为二元随机变量，我们分别给出以下定义：

(1) Y关于X的条件熵：

$$H(Y/X)=H(X,\ Y)-H(X)=\sum_{x}\sum_{y}p(x,y)\log q(y/x); \tag{1.8}$$

(2) X关于Y的条件熵：

$$H(X/Y)=H(X,\ Y)-H(X)=\sum_{x}\sum_{y}p(x,y)\log p(x/y); \tag{1.9}$$

(3) X与Y的交互信息：

$$I(X;\ Y)=H(X)+H(Y)-H(X,\ Y)$$

$$=\sum_{x}\sum_{y}p(xy)\log\frac{p(x,\ y)}{p(x)q(y)}, \tag{1.10}$$

其中

$$p(x/y)=p(x,\ y)/q(y),\ q(y/x)=p(x,y)/p(x).$$

条件熵给出了两个随机变量的条件不肯定性，也就是，在一个随机变量固定的条件下另一个随机变量的条件不肯定性。而交互信息则反映了两个随机变量的相互依赖程度。

引理2 (1) 对条件熵有不等式

$$0\leqslant H(Y/X)\leqslant H(Y), \tag{1.11}$$

其中$H(Y/X)=0$的充要条件是Y由X确定，而$H(Y/X)=H(Y)$的充要条件是X与Y相互独立。

(2) 对交互信息$I(X;\ Y)$有不等式

$$0\leqslant I(X;\ Y)\leqslant\min\{H(X),\ H(Y)\}, \tag{1.12}$$

其中$I(X;\ Y)=0$的充要条件是X与Y相互独立，而$I(X;\ Y)=\min\{H(X),\ H(Y)\}$的充要条件是$Y$由$X$确定或$X$由$Y$确定。

证明：首先由不等式(1.5) 可得$I(X;\ Y)\geqslant 0$成立，且等号成立的充要条件是X与Y相互独立。因为

$$\begin{aligned}I(X;\ Y)&=H(X)+H(Y)-U(X,\ Y)\\&=H(Y)-H(Y/X),\end{aligned}$$

所以有 $H(Y)\geqslant H(Y/X)$成立，且等号成立的充要条件是X与Y相互独立。

同样由(1.5)式可得$H(Y/X)=H(X,Y)-H(X)\geqslant 0$成立，且等号成立的充要条件是$Y$由$X$确定。这由交互信息的定义可得

$$I(X;\ Y)=H(Y)-H(Y/X)\leqslant H(Y),$$

且$I(X;\ Y)=H(Y)$的充要条件是Y由X确定。引理得证。

从引理2的不等式可以看到，交互信息反映了两个随机变量的相互依赖程度，它比概率论中的相关系数具有更强的特征。例如，对交互信息的“$I(X;\ Y)=0$的充要条件是X与Y相互独

立"的性质，说明了交互信息比相关系数更能反映两个随机变量的依赖程度．相关系数为零只是只给出了两个随机变量X与Y相互独立的一个充分条件．因此，在近代非线性统计分析中，常把交互信息作为两个随机变量的依赖度．

3．信息量的集合关系表示

如果把以上定义的仙农熵、联合熵、条件熵与交互信息都称为信息量，那么它们的相互关系与集合的并、差、交关系相似．如图3.1所示，如果把$H(X)$，$H(Y)$看作基集，那么相应的联合熵、条件熵与交互信息就为基集的并、差、交．这样我们可以用以下关系表示．

信息量的并：$H(X\vee Y)=H(X,\ Y)$；

信息量的差：$H(X\setminus Y)=H(X/Y)$，

$H(Y\setminus X)=H(Y/X)$；

信息量的交：$H(X\wedge Y)=I(X;\ Y)$．

以上讨论可直接推广到多元随机变量的情形中去．也就是，除了由上述并、差、交运算产生的信息量之外，还可由它们的混合运算产生各种新的随机变量．例如，由三个随机变量X、Y、Z的基本信息集$H(X)$、$H(Y)$与$H(Z)$的交差混合运算可产生条件交互信息：

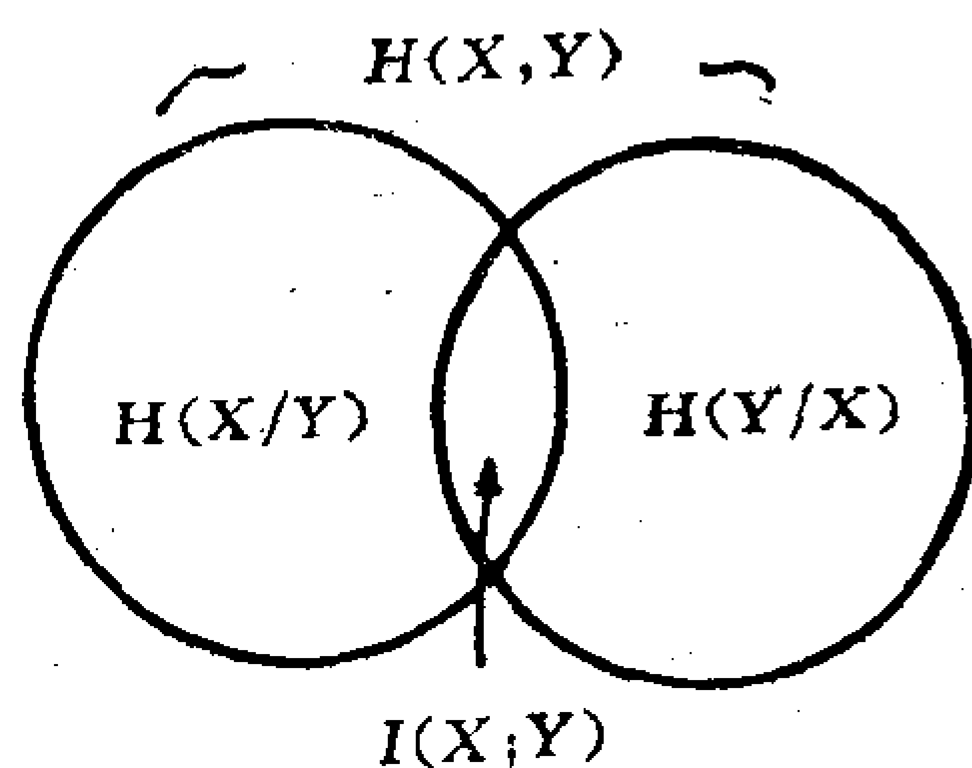


图3.1 信息量的集合关系表示

$$I(X;\ Y/Z)=H[(X\wedge Y)\setminus Z]$$

$$=\sum_x\sum_y\sum_z p(x,y,z)\log\frac{p(x,y/z)}{p(x/z)p(y/z)}. \qquad (1.13)$$

在(1.13)中，$p(x,y,z)$是(X,Y,Z)的联合分布，而相应的条件

分布、边际分布都可由$p(x,y,z)$确定如下：

$$p(x,y/z)=\frac{p(x,y,z)}{r(z)},\quad p(x/z)=\frac{p(x,z)}{r(z)},$$

$$q(y/z)=\frac{q(y,\ z)}{r(z)}.$$

其中

$$p(x,z)=\sum_{y}p(x,y,z),\quad q(y,z)=\sum_{x}p(x,y,z),$$

$$r(z)=\sum_{x}p(x,z)=\sum_{y}q(y,z).$$

由三个随机变量X、Y、Z的基本信息集$H(X)$、$H(Y)$与$H(Z)$的并、差、交混合运算产生的各种信息量如图3.2所示。其中每一块子集均可代表一种信息量。如

集合Ⅰ表示$H[X\setminus(Y\vee Z)]$；

集合Ⅳ表示$H[(X\wedge Y)\setminus Z)]$；

集合Ⅰ，Ⅱ，Ⅳ的并表示$H[(X\vee Y)\setminus Z)]$.

对此就不一一列举了。需要说明的一点是这些信息量都可用类似于(1.13)的关系式表示，它们由联合分布$p(x,y,z)$完全确定.

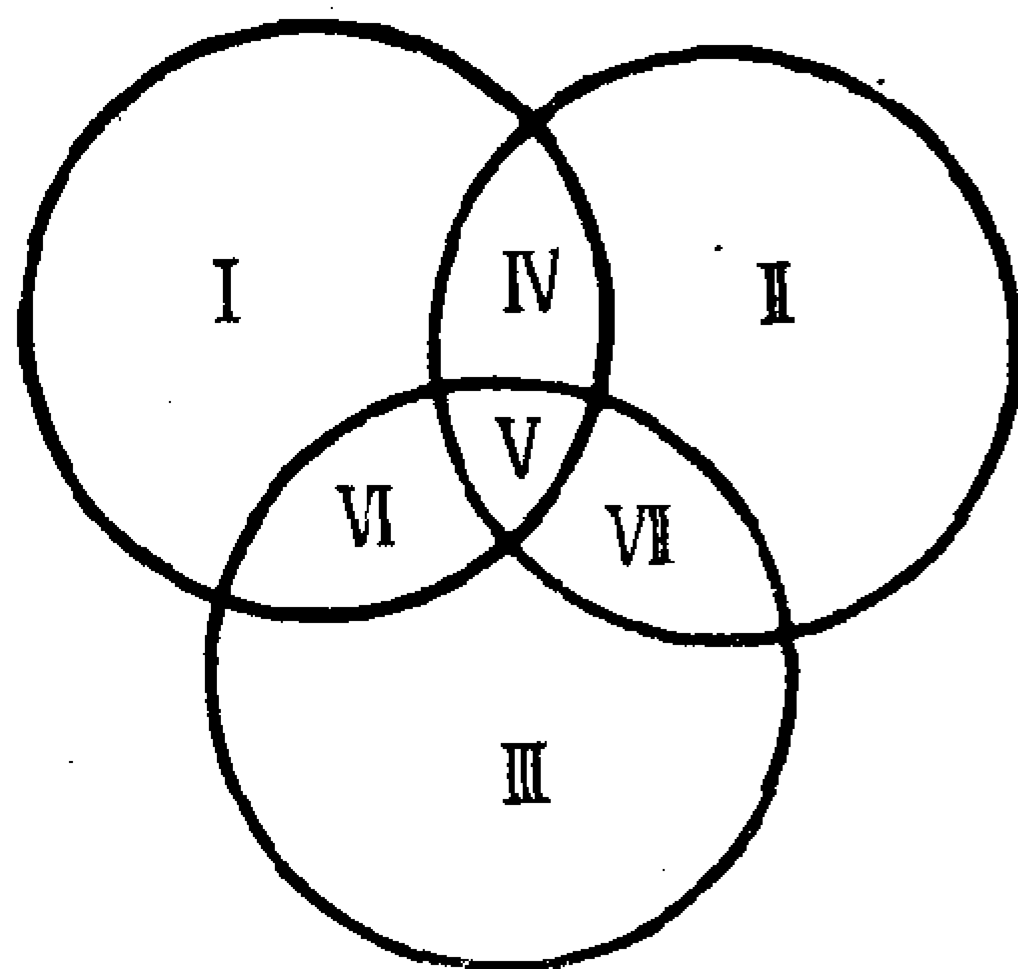


图3.2 三个随机变量的信息量的集合关系

4.条件交互信息的性质

利用概率论的分布性质可得到条件交互信息的各种性质，对此概述如下：

(1) 不等式(1.12)可推广到条件交互信息情形，这时

$$0 \leqslant I(X;\ Y/Z) \leqslant \min\{H(X/Z),\ H(Y/Z)\},\quad (1.14)$$

且$I(X;\ Y/Z)=0$的充要条件是$(X,\ Y)$关于Z条件独立，也就是有

$$p(x,y/z)=p(x/z)q(y/z)$$

成立。而$I(X;\ Y/Z)=\min\{H(X/Z),\ H(Y/Z)\}$的充要条件是$X$由$(Y,\ Z)$确定或$Y$由$(X,\ Z)$确定。

(2) 如果$f(y)$是一个从$\mathbf{Y}\to\mathbf{Z}$的单值映射，那么必有

$$I(X;\ Y) \geqslant I(X;\ f(Y)) \quad (1.15)$$

成立，且等号成立的充要条件是$f(y)$是一个1—1映射。由(1.15)即可推出对任何随机变量X，Y，Z有

$$I[X;(Y,\ Z)] \geqslant I(X;\ Y) \quad (1.16)$$

成立，且等号成立的充要条件是$X\to Y\to Z$是一个马氏链(Markov链)，也就是，$(X,\ Z)$ 关于Y条件独立。

(3) 柯尔莫各洛夫 (Kolmogorov)公式。对任何随机变量X，Y，Z总有关系式

$$I[X;(Y,Z)]=I(X;\ Y)+I(X;\ Z/Y) \quad (1.17)$$

成立。

定理中，(1)、(2)的证明与引理2相同，性质(3)的证明由概率分布的性质可得，对此可见[26]文详述。

§3.2 信道容量及其性质

1.信道容量的定义

信道与信道编码的定义在§2.1中已经给定。如§2.1所记，设

$$\boldsymbol{C}=[\boldsymbol{U},\ p(v/u),\ \boldsymbol{V}] \tag{2.1}$$

为信道。我们称$\boldsymbol{U}$上的概率分布$\boldsymbol{S}_u=[\boldsymbol{U},\ p(u)]$为信道的入口分布。如果信道$\boldsymbol{C}$与信道入口分布$\boldsymbol{S}_u$给定，那么信道的输入、输出信号的概率分布确定为

$$p(u,\ v)=p(u)\cdot p(v/u),\ u\in\boldsymbol{U},\ v\in\boldsymbol{V}, \tag{2.2}$$

我们记U、V为相应的信道输入、输出信号随机变量,具有联合分布

$$P_r\{(U,\ V)=(u,\ v)\}=p(u,\ v).$$

这时（U, V）的交互信息为

$$I(U;\ V)=\sum_u\sum_v p(u,\ v)\log\frac{p(u,\ v)}{p(u)q(v)}, \tag{2.3}$$

其中$p(u)$、$q(v)$分别为随机变量U、V的概率分布。因为（U, V)的联合分布由信道$\boldsymbol{C}$与入口分布$\boldsymbol{S}_u$确定,因此我们又记(2.3)的交互信息为$I(U;\ V)=I(\boldsymbol{S}_u;\ \boldsymbol{C})$。

定义1 对已给信道$\boldsymbol{C}=[\boldsymbol{U},\ p(v/u),\ V]$，我们称

$$C(\boldsymbol{C})=\sup\{I(\boldsymbol{S}_u;\ \boldsymbol{C}),\ \boldsymbol{S}_u\in\boldsymbol{S}(\boldsymbol{U})\} \tag{2.4}$$

为信道$\boldsymbol{C}$的信道容量，其中$\boldsymbol{S}(\boldsymbol{U})$为全体$\boldsymbol{U}$上的概率分布。

因为$\boldsymbol{S}(\boldsymbol{U})=\{(p(u),u\in\boldsymbol{U}):\ p(u)\geqslant 0,\ \sum_u p(u)=1\}$是一个凸闭集，而$I(\boldsymbol{S}_u,\boldsymbol{C})$是$(p(u),\ u\in\boldsymbol{U})$的连续函数,因此有

$$C(\boldsymbol{C})=\max\{I(\boldsymbol{S}_u;\ \boldsymbol{C}),\ S_u\in\boldsymbol{S}(\boldsymbol{U})\} \tag{2.5}$$

成立。这时，有一个入口分布$\boldsymbol{S}_{u,0}=[\boldsymbol{U},\ p_0(u)]\in\boldsymbol{S}(\boldsymbol{U})$,使

$$I(\boldsymbol{S}_{u,0},\ \boldsymbol{C})=C(\boldsymbol{C})=\max\{I[\boldsymbol{S}_u;\boldsymbol{C}],\ \boldsymbol{S}_u\in\boldsymbol{S}(\boldsymbol{U})\} \tag{2.6}$$

成立。

对已给多重信道$\boldsymbol{C}^n=[\boldsymbol{U}^n,p(v^n/u^n),\ \boldsymbol{V}^n]$，我们同样可引进它的信道容量

$$C(\boldsymbol{C}^n)=\max\{I(\boldsymbol{S}_u^n;\boldsymbol{C}^n),\ \boldsymbol{S}_u^n\in\boldsymbol{S}(\boldsymbol{U}^n)\}\qquad(2.5')$$

其中$\boldsymbol{S}(\boldsymbol{U}^n)$为全体$\boldsymbol{U}^n$上的概率分布，相应的元为$\boldsymbol{S}_u^n$，这时

$$I(U^n;\ V^n)=\sum_{u^n}\sum_{v^n}p(u^n,\ v^n)\log\frac{p(u^n,\ v^n)}{p(u^n)q(v^n)},\qquad(2.3')$$

其中$I(U^n;\ V^n)=I(\boldsymbol{S}_u^n;\boldsymbol{C}^n)$，而$p(u^n)$，$q(v^n)$与(2.3)的$p(u)$、$q(v)$类似定义。

2.无记忆信道的信道容量

对已给多重信道序列

$$\boldsymbol{C}^I=\{\boldsymbol{C}^n=[\boldsymbol{U}^n,\ p(v^n/u^n),\ \boldsymbol{V}^n],\ n=1,2,3\cdots\},\qquad(2.7)$$

如果

$$u^n=(u_1,\cdots,u_n),\ u_i\in U,\ v^n=(v_1,\cdots,v_n),\ v_i\in\boldsymbol{V},$$

而且

$$p(v^n/u^n)=\prod_{i=1}^{n}p(v_i/u_i),\qquad(2.8)$$

那么，我们称$\boldsymbol{C}^n$是由$\boldsymbol{C}=[\boldsymbol{U},\ p(v/u),\ \boldsymbol{V}]$产生的$n$-重无记忆信道。

定理2 如果$\boldsymbol{C}^n$是由$\boldsymbol{C}$产生的n-重无记忆信道，那么有

$$C(\boldsymbol{C}^n)=n\cdot C(\boldsymbol{C})\qquad(2.9)$$

成立，其中$C(\boldsymbol{C}^n)$，$C(\boldsymbol{C})$分别为信道$\boldsymbol{C}^n$与$\boldsymbol{C}$的信道容量.

证明：对n-重无记忆信道$\boldsymbol{C}^n$，如果$p(u^n)$为任一入口分布，如果$p(u^n)$给定，那么信道$\boldsymbol{C}^n$的输入、输出随机变量(U^n,V^n)确定。这时

$$I(U^n,\ V^n)=H(V^n)-H(V^n/U^n)\qquad(2.10)$$

为随机变量$(U^n,\ V^n)$的交互信息。如果记

$$U^n=(U_1, \cdots, U_n),\ V^n=(V_1, \cdots, V_n),$$

那么由联合熵的性质可得

$$H(V^n)\leqslant\sum_{i=1}^{n}H(V_i) \tag{2.11}$$

成立。由 (2.8)式可得

$$\begin{aligned}H(V^n/U^n)&=\sum_{u^n}\sum_{v^n}p(u^n, v^n)\log p(v^n/u^n)\\&=\sum_{u^n}\sum_{v^n}p(u^n, v^n)\left[\sum_{i=1}^{n}\log p(v_i/u_i)\right]\\&=\sum_{i=1}^{n}H(V_i/U_i).\end{aligned} \tag{2.12}$$

将(2.11)、(2.12)代入(2.10)即得

$$\begin{aligned}I(U^n, V^n)&\leqslant\sum_{i=1}^{n}[H(V_i)-H(V_i/U_i)]\\&=\sum_{i=1}^{n}I(U_i; V_i)\leqslant n\cdot C(\mathbf{C})\end{aligned} \tag{2.13}$$

成立。因为$p(u^n)$是$\mathbf{C}^n$的任意入口分布，所以必有$C(\mathbf{C}^n)\leqslant n\cdot C(\mathbf{C})$成立。

另一方面，如我们取

$$p_0(u^n)=p_0(u_1)\times p_0(u_2)\times\cdots\times p_0(u_n) \tag{2.14}$$

为$\mathbf{C}^n$的入口分布，其中$p_0(u)$为使 (2.6) 成立的入口分布。这时相应的联合分布为

$$p_0(u^n, v^n)=p_0(u^n)p(v^n/u^n)=\prod_{i=1}^{n}p_0(u_i, v_i), \tag{2.15}$$

其中$p_0(u, v)=p_0(u)p(v/u)$．如果我们记$I_0(U^n; V^n)$，$I_0(U; V)$分别为由联合分布$p_0(u^n, v^n)$，$p_0(u, v)$决定的交互信息，那么由(2.15)即得

$$I_0(U^n; V^n)=\sum_{i=1}^{n} I_0(U, V)=n\cdot C(\mathbf{C})$$

成立，由信道容量的定义可知$I_0(U^n; V^n)\leqslant C(\mathbf{C}^n)$必成立，因此有$n\cdot C(\mathbf{C})\leqslant C(\mathbf{C}^n)$成立．结合本定理的第一部分证明即得$n\cdot C(\mathbf{C})=C(\mathbf{C}^n)$成立．定理得证．

3.若干典型信道及其容量

由信道容量的定义可知，当信道$\mathbf{C}$给定时，它的容量$C(\mathbf{C})$是唯一确定的，在一般情形，$C(\mathbf{C})$可由$I[\mathbf{S}_u; \mathbf{C}]$函数对入口分布$p(u)$求条件极值而得．以下我们讨论几种在通信中常见的典型信道及它们的信道容量．它们是二进对称信道、Z-信道与M-信道，它们的构造如图3.3所示．

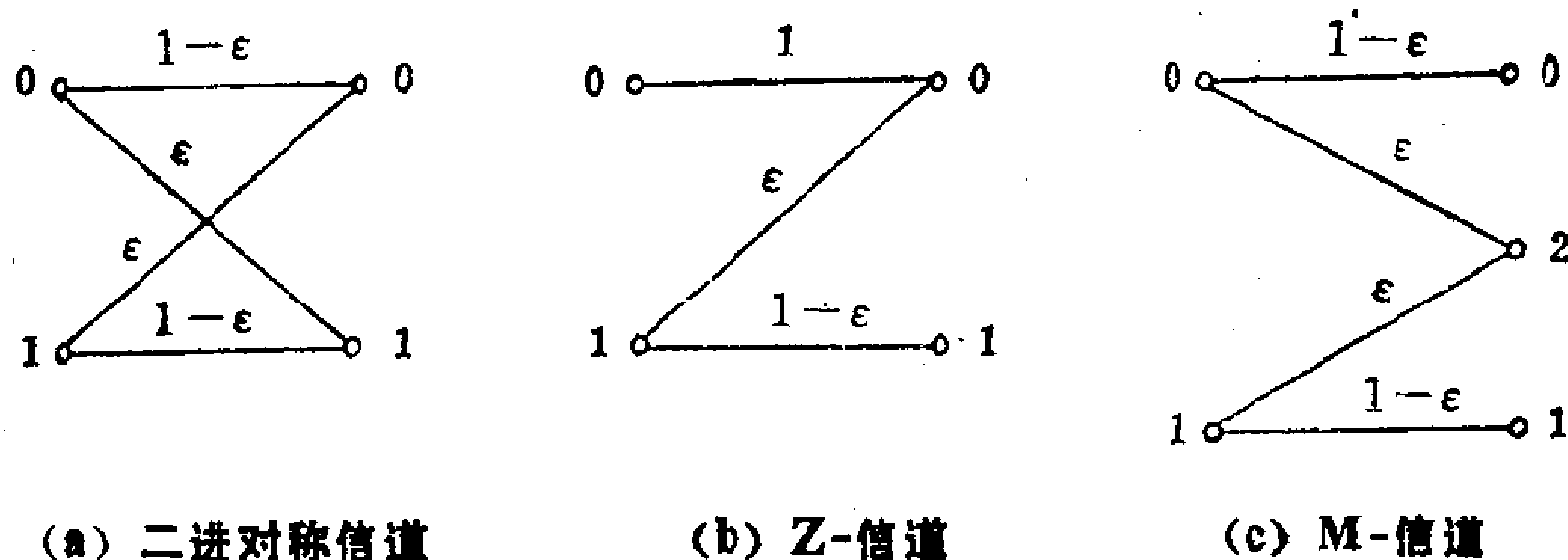


图3.3

对这三种信道，我们描述它们的信道转移概率(条件概率)与计算它们的联合概率分布、交互信息与信道容量．因为这三种信道的输入信号字母表都是二元集合$\mathbf{U}=\{0, 1\}$，因此我们记它们的入口分布为

$$p(1)=p,\ p(0)=q=1-p,\ 0\leqslant p\leqslant 1. \tag{2.16}$$

(1) 二进对称信道(不失一般性，取p，$\varepsilon\leqslant 0.5$)。

转移概率：$p(0/0)=p(1/1)=1-\varepsilon$,

$$p(1/0)=p(0/1)=\varepsilon; \tag{2.17}$$

联合概率分布：$p(0,\ 0)=p(1,\ 1)=(1-\varepsilon)p$,

$$p(1,\ 0)=p(0,\ 1)=q\varepsilon; \tag{2.18}$$

交互信息：$I(U;\ V)=H(p+\varepsilon-2p\varepsilon)-H(\varepsilon)$; (2.19)

信道容量：$C(\boldsymbol{C}_1)=1-H(\varepsilon)$, (2.20)

其中 $H(\delta)\equiv-\delta\log\delta-(1-\delta)\log(1-\delta)$.

(2) Z-信道.

转移概率：$p(0/0)=1$，$p(1/0)=0$，$p(1/0)=1-\varepsilon$,

$$p(0/1)=\varepsilon; \tag{2.17'}$$

联合概率分布：$p(0,\ 0)=p+q\varepsilon$，$p(0,\ 1)=0$,

$$p(1,\ 0)=q\varepsilon, p(1,\ 1)=q(1-\varepsilon); \tag{2.18'}$$

交互信息：$I(U;\ V)=H(p+q\varepsilon)-q\cdot H(\varepsilon)$; (2.19′)

信道容量：$C(\boldsymbol{C}_2)=H(p_0+q_0\varepsilon)-q_0\cdot H(\varepsilon)$, (2.20′)

其中$q_0=1-p_0$，而p_0为方程

$$(1-\varepsilon)\log[r(p,\ \varepsilon)]-H(\varepsilon)=0, \tag{2.21}$$

之解，其中$r(p,\ \varepsilon)=[(1-\varepsilon)q]^{-1}-1$.

(3) M-信道(不失一般性，取$p\leqslant 0.5$)。

转移概率：$p(0/0)=p(1/1)=1-\varepsilon$,

$p(1/0)=p(0/1)=0$,

$$p(2/0)=p(2/1)=\varepsilon; \tag{2.17''}$$

联合概率分布：$p(0,\ 0)=p(1,\ 1)=0$，$p(0,\ 2)=p\varepsilon$,

$p(1,\ 2)=q\varepsilon$

$p(0,\ 0)=(1-\varepsilon)p$，$p(1,\ 1)=(1-\varepsilon)q$;

(2.18″)

交互信息：$I(U;\ V)=(1-\varepsilon)H(p)$；　　(2.19″)

信道容量，$C(\boldsymbol{C}_3)=1-\varepsilon$。　　(2.20″)

§3.3 无记忆信道的编码定理

1.信道序列的编码问题

形象地说，所谓信道序列的编码问题就是有多少信息可以通过给定的信道．在一个实际的通信问题中，信道的客观物理条件（如通信设备、线路情况、传递的消息范围及其概率分布情况等）是完全确定的，但是其中信号的结构形式是可以作适当调整的．所谓信息在信道中通过一般指这些信息能正确的经过信道无误差地传给对方．因此，信道序列的编码问题就是指对固定的通信序列

$$\boldsymbol{E}^n=\{\boldsymbol{S}^n,\ \boldsymbol{C}^n\}=\{[\boldsymbol{X}^n,\ p(x^n)],\ [\boldsymbol{U}^n,p(v^n/u^n),\boldsymbol{V}^n]\}$$

$$n=1,2,3,\cdots,\qquad (3.1)$$

寻找适当的编码函数$(f^n,\ g^n)$，它们分别为

$$f^n:\ \boldsymbol{X}^n\to\boldsymbol{U}^n;\ g^n:\ \boldsymbol{V}^n\to\boldsymbol{Y}^n\qquad (3.2)$$

的映射．由§2.1的讨论可知，如果$\boldsymbol{E}^n$，$(f^n,\ g^n)$给定，那么通信系统的输入、输出消息与信号的随机变量$(X^n,\ U^n,\ V^n,\ Y^n)$确定，它们的联合分布为

$$\begin{aligned}&p(x^n,\ u^n,\ v^n,\ y^n)\\&=p(x^n)f^n(u^n/x^n)p(v^n/u^n)g^n(y^n/v^n),\end{aligned}\qquad (3.3)$$

其中

$$f^n(u^n/x^n)=\begin{cases}1,\ \text{如}u^n=f^n(x^n),\\0,\ \text{否则},\end{cases}$$

$$g^n(y^n/v^n)=\begin{cases}1, & 如 y^n=g^n(v^n)\\ 0, & 否则.\end{cases} \tag{3.4}$$

在本节中，我们取

$$\boldsymbol{X}^n=\boldsymbol{Y}^n=\{1,2,\cdots,M^n\}, \tag{3.5}$$

$$p(x^n)=1/M^n, \ 对任何 x^n\in\boldsymbol{X}^n. \tag{3.6}$$

定义1 称R为信道序列$\boldsymbol{C}^n$的一个可达速率，如果存在一个由(3.5)、(3.6)给定的信源序列$\boldsymbol{S}^n$，对任何$\varepsilon>0$，只要n充分大，就有

$$M^n\geqslant\exp_2[n\cdot R\cdot(1-\varepsilon)], \tag{3.7}$$

且有一列适当的编码函数(f^n, g^n)，使得平均误差概率为

$$p_e(f^n, g^n)=P_r\{X^n\neq Y^n\}\leqslant\varepsilon, \tag{3.8}$$

其中X^n，U^n，V^n，Y^n为由$\boldsymbol{S}^n$，$\boldsymbol{C}^n$，(f^n, g^n)确定的随机变量。全体可达速率的上确界R_0为最大可达速率，这时对任何$R'>R_0$，总有一个$\varepsilon_0>0$，对任何$N>0$，总有一个$n>N$，只要

$$M^n\geqslant\exp_2[n\cdot R_0\cdot(1-\varepsilon_0)], \tag{3.8'}$$

那么对任何编码函数(f^n, g^n)，必有$p_e(f^n, g^n)>\varepsilon_0$成立。

信道序列的编码问题就是对已给的信道序列$\boldsymbol{C}^n$，$n=1,2,\cdots$，求它的最大可达速率。在有些文献中，也把最大可达速率称为信道容量。为了区别起见，在本书中我们给以不同的称呼，它们在一定条件下相等。

2.信道序列的正编码定理

在本段中，我们先讨论信道序列的正编码定理，也就是找出有关信道的可达速率。

定理1 （无记忆信道序列的正编码定理）。如果$\boldsymbol{C}^n$为由$\boldsymbol{C}$生成的无记忆信道序列，那么信道$\boldsymbol{C}$的容量$C(\boldsymbol{C})$为信道序列$\boldsymbol{C}^n$的一个可达速率。

证明：对这个定理的证明方法有多种。在本书中我们采用随机码的方法来证，这个方法比较巧妙地运用了概率论的工具，因此比较简单，且在其它编码理论中还有应用。对本定理的证明要点如下：

(1) 随机编码。设$\boldsymbol{S}^n$为由(3.5)、(3.6)给定的信源序列，记$\boldsymbol{C}^n$为由(2.10)给定的无记忆信道，而

$$\boldsymbol{S}_{u,0}=[\boldsymbol{U},p_0(u)],\quad \boldsymbol{S}_{u,0}^n=[\boldsymbol{U}^n,p_0(u^n)],\quad(3.9)$$

分别为信道$\boldsymbol{C}$，$\boldsymbol{C}^n$的入口分布，由(2.6)，(2.15)给定。以下记

$$\boldsymbol{U}_0^n=\{U_1^n,\ U_2^n,\ \cdots,\ U_{m^*}^n\}\qquad(3.10)$$

为一组独立同分布的随机向量，每个

$$U_i^n=(U_{i,1},\ U_{i,2},\ \cdots,\ U_{i,n})\qquad(3.11)$$

为在$\boldsymbol{U}^n$中取值，具概率分布$p_0(u^n)$。记$\boldsymbol{U}_0^n$的样本值为

$$u_0^n=\{u_1^n,\ u_2^n,\ \cdots,\ u_{m^*}^n\}。\qquad(3.10')$$

定义1 一个随机编码f^{n*}是一个从$\boldsymbol{X}^n=\{1,2,\ \cdots,\ M^n\}$到$\boldsymbol{U}_0^n=\{U_1^n,\ U_2^n,\ \cdots,\ U_{m^*}^n\}$的映射。这时

$$f^{n*}(i)=U_i^n,\quad i=1,\ 2,\ \cdots,\ M^n。\qquad(3.12)$$

一个随机编码f^{n*}的样本值是一个从$\boldsymbol{X}^n$到$\boldsymbol{u}_0^n$的映射f^n，这时

$$f^n(i)=u_i^n,\quad i=1,2,\cdots,M^n\qquad(3.12')$$

(2) 信息门限译码函数。

对一个(3.12′)的编码$f^n(i)=u_i^n$，如果$v^n=(v_1,\cdots,\ v_n)$是一个信道$\boldsymbol{C}^n$的接收信号向量。我们构造一个译码函数g^n是一个从$\boldsymbol{V}^n$到$\boldsymbol{Y}^n=\boldsymbol{X}^n$的映射，由

$$g^n(v^n)=\begin{cases}j,\ 如果i_0(u_j^n;\ v^n)\geq K^n,\\ \boldsymbol{Y}^n中任取一值，否则\end{cases}\qquad(3.13)$$

定义，其中K^n为一个适当的门限值，它的取值在下文中给出，

而$i_0(u^n; v^n)$为联合分布$p_0(u^n, v^n)$的交互信息密度，由

$$i_0(u^n, v^n)=\log\frac{p_0(u^n, v^n)}{p_0(u^n)q_0(v^n)} \tag{3.14}$$

定义，其中$q_0(v^n)=\sum_{u^n} p_0(u^n, v^n)$。我们称这个译码函数为信息门限译码函数。

(3) 随机码的平均误差概率。

在本节2-(1)，(2)段的讨论中，我们给出了信源、信道序列$\mathbf{S}^n$，$\mathbf{C}^n$与编码函数（f^n，g^n）的定义。因此，相应的误差概率$p_e(f^n, g^n)$确定。所谓随机码的平均误差概率就是指在编码函数f^n为随机编码f^{n*}时相应误差概率的平均值，我们记之为$p_{AVE}(f^{n*}, g^n)$。

为了估计$p_{AVE}(f^{n*}, g^n)$的值，对固定的发送消息i，我们记相应的发送信号与接收信号为U_i^n，V_i^n，这时（U_i^n，V_i^n)的联合概率分布为

$$P_r\{(U_i^n, V_i^n)=(u^n, v^n)\}=p_0(u^n, v^n), \tag{3.15}$$

其中$p_0(u^n, v^n)$由（2.17）定义。

对$j\neq i$的情形，我们记U_j^n为发送消息j时的信道输入信号，而V_i^n为发送消息i的接收信号。这时U_j^n与V_i^n的二个相互独立的随机变量，它们的联合概率分布为

$$p_1(u^n, v^n)=P_r\{U_j^n, V_i^n)=(u^n, v^n)\}=p_0(u^n)q_0(v^n), \tag{3.15'}$$

其中$p_0(u^n, v^n)$由（2.16）定义，而$q_0(v^n)=\sum_{u^n} p_0(u^n, v^n)$。

为了估计 $p_{AVE}(f^{n*}, g^n)$ 的值，我们还必需分析产生误差$p_{AVE}(f^{n*},g^n)$的原因，对此定义产生$p_{AVE}(f^{n*},g^n)$的两类误差。以下记

$$P_{I,i}(f^{n*},\ g^n)=P_r\{\text{输入消息}i\text{时},\ g^n(V_i^n)\neq i\},\tag{3.16}$$

$$P_{II;i}(f^{n*},\ g^n)=P_r\{\text{输入消息}i\text{时},\ \text{有一个}j\neq i,\ \text{使}g^n(V_j^n)=i\}.\tag{3.16'}$$

我们分别称

$$p_a(f^{n*},\ g^n)=\frac{1}{M^n}\sum_{i=1}^{n}p_{a,i}(f^{n*},g^n),a=I,II,\tag{3.17}$$

为随机编码的第一、二类平均误差．这时有

$$p_{AVE}(f^{n*},\ g^n)\leqslant p_I(f^{n*},g^n)+p_{II}(f^{n*},g^n)\tag{3.18}$$

成立．另一方面，又由随机编码的概率分布的对称性可得

$$p_I(f^{n*},\ g^n)=p_{I,i}(f^{n*},g^n),$$

$$p_{II}(f^{n*},g^n)=p_{II,i}(f^{n*},\ g^n)$$

对任何$i=1,2,\ \cdots,\ M^n$成立．因此，在下文中，我们只要估计$p_{I,1}(f^{n*},\ g^n)$与$p_{II;1}(f^{n*},\ g^n)$的值就可．

(4) $p_{I,1}(f^{n*},\ g^n)$与$p_{II,1}(f^{n*},\ g^n)$的估计．

为估计$p_{I,1}(f^{n*},\ g^n)$与$p_{II,1}(f^{n*},g^n)$的值，对任何$\varepsilon>0$，我们首先取$M^n=\exp_2[n\cdot R\cdot(1-\varepsilon)]$，其中$R=C(\mathbf{C})$为信道$\mathbf{C}$的信道容量，而选取(3.13)的门限值为$K^n=n\cdot R\cdot(1-\varepsilon/2)$．这时对第一类误差的估计式为

$$p_{I;1}(f^{n*},\ g^n)=P_r\{i_0(U_1^n,\ V_1^n)<K^n\}.\tag{3.19}$$

因为$(U_1^n,\ V_1^n)$的联合分布$p_0(u^n,\ v^n)$是(2.17)的乘积分布，因此它们的分量$(U_{1;k},V_{1;k})$是独立同分布的随机变量，且联合分布为(2.17)中定义的$p_0(u,\ v)$．这时，由大数定律可得

$$\frac{1}{n}i_0(u_1^n,\ v_1^n)=\frac{1}{n}\sum_{k=1}^{n}i_0(u_{1,k},v_{1,k})\to\boldsymbol{E}\{i_0(U;\ V)\}$$
$$=C(\mathbf{C})=R,\tag{3.20}$$

其中“→”表示以概率收敛的极限，而$\boldsymbol{E}\{Z\}$为随机变量Z的数学期望。由关系式(3.19)、(3.20)即得，对任何$\varepsilon>0$，只要n充分大，就有

$$p_I(f^{n*},g^n)=p_{I;\ 1}(f^{n*},\ g^n)<\varepsilon/2 \tag{3.21}$$

成立。对第二类误差概率的估计有

$$\begin{aligned} &p_{II;\ 1}(f^{n*},\ g^n)\\ &=P_r\{\text{有一个}j\neq 1,\ \text{使}i_0(U_j^n;\ V_1^n)\geqslant K^n\}\\ &\leqslant(M^n-1)\cdot P_r\{i_0(U_2^n;\ V_1^n)\geqslant K^n\} \end{aligned} \tag{3.22}$$

成立，其中不等式由$(U_j^n,\ V_1^n)$，$j\neq 1$，的同分布性质而得。因为$(U_j^n,\ V_1^n)$的联合概率分布是$p_1(u^n,\ v^n)$由(3.15′)定义。如我们记

$$\boldsymbol{A}_0^n=\boldsymbol{A}_0^n(K^n)=\{(u^n,\ v^n);\ i_0(u^n,\ v^n)\geqslant K^n\}. \tag{3.23}$$

那么，对任何$(u^n,\ v^n)\in\boldsymbol{A}_0^n$，总有

$$i_0(u^n,\ v^n)=\log\{p_0(u^n,\ v^n)/[p_0(u^n)q_0(v^n)]\}\geqslant K^n$$

成立，此即有

$$p_1(u^n,\ v^n)=p_0(u^n)q_0(v^n)\leqslant p_0(u^n,\ v^n)\exp_2(-K^n) \tag{3.24}$$

成立。因为对任何$(u^n,\ v^n)\in\boldsymbol{A}_0^n$总有(3.24)成立，因此必有

$$p_1(\boldsymbol{A}_0^n)\leqslant p_0(\boldsymbol{A}_0^n)\exp_2(-K^n)\leqslant\exp_2(-K^n) \tag{3.25}$$

成立。由(3.22)，(3.25)与M^n，K^n的取值可得

$$\begin{aligned} &p_{II;\ 1}(f^{n*},\ g^n)\leqslant(M^n-1)\cdot p_1(\boldsymbol{A}_0^n)\\ &\leqslant\exp_2[n\cdot R\cdot(1-\varepsilon)+K^n]=\exp_2(-n\cdot\varepsilon/2)\to 0, \end{aligned}$$

当$n\to\infty$时。这样，当n充分大时，就有

$$p_{II}(f^{n*},\ g^n)=p_{II;\ 1}(f^{n*},\ g^n)<\varepsilon/2 \tag{3.26}$$

成立。由关系式(3.18)，(3.21)与(3.26)即得$p_{AVE}(f^{n*},\ g^n)\leqslant\varepsilon$。此即$R=C(\boldsymbol{C})$为可达速率。定理得证。

3.一般信道序列的反编码定理

在定理1中，我们给出了无记忆信道序列的正编码定理，该定理确立了无记忆信道的信道容量$C(\mathbf{C})$是一个可达速率．我们现在讨论它的反编码定理，也就是证明这个信道容量$C(\mathbf{C})$是无记忆信道序列的最大可达速率．为此我们先给出一个常用的不等式．

引理1 （法诺(Fano)不等式）．设X、Y是两个随机变量，在$\boldsymbol{X}=\boldsymbol{Y}$上取值．如记$p_e=P_r\{X\neq Y\}$，那么有

$$H(X/Y)\leqslant H(p_e)+p_e\cdot\log(\|\boldsymbol{X}\|-1),\quad(3.27)$$

其中$H(X/Y)$是条件熵，$H(p)=-p\cdot\log p-(1-p)\cdot\log(1-p)$．

证明：如记$p(x,\ y)$是随机变量$(X,\ Y)$的联合概率分布，而

$$\boldsymbol{A}_0=\{(x,\ y):\ x=y\},\quad \boldsymbol{A}_1=\{(x,\ y):\ x\neq y\}$$

是$\boldsymbol{X}\times\boldsymbol{Y}$的两个子集．这时有$p(\boldsymbol{A}_0)=1-p_e$，$p(\boldsymbol{A}_1)=p_e$．

如我们定义$Z=f(X,Y)$为一个由(X,Y)确定的随机变量，其中$f(x,\ y)=0$或1，分别当$(x,\ y)\in\boldsymbol{A}_0$或$\boldsymbol{A}_1$时．这时$(X,\ Y,\ Z)$的联合概率分布为

$$p(x,\ y,\ z)=\begin{cases}p(x,\ y), & \text{如}z=f(x,\ y)\text{时},\\ 0, & \text{否则}.\end{cases}$$

由条件熵的定义可得

$$\begin{aligned}H(X/Y)&=-\sum_{x,\ y,\ z}p(x,\ y,\ z)\log p(x/y)\\&=\sum_{z}r(z)\sum_{x,\ y}[p(x,yz)/r(z)]\\&\quad\cdot\log[p(x/y)]^{-1},\end{aligned}$$

其中$r(z)=\sum\limits_{x,\ y}p(x,y,z)$．这时有$\sum\limits_{x,\ y}[p(x,y,z)/r(z)]=1$ 成

立，而且有$r(0)=1-p_e, r(1)=p_e$。因为$f(p)=\log p$在p的定义域中是一个上凸函数，因此有

$$H(X/Y)\leqslant\sum_{z} r(z)\cdot\log\{r(z)^{-1}\cdot\sum_{x,y}[p(x,y,z)/r(z)]\}$$

$$=-\sum_{z} r(z)\log r(z)+\sum_{z} r(z)\cdot\log\sum_{x,y}[p(x,y,z)/r(z)]$$

$$=H(p_e)+p_e\cdot\log(\|\boldsymbol{X}\|-1)\qquad(3.28)$$

成立。因为这时有

$$\sum_{x,y}[p(x,y,1)/p(x/y)]$$

$$=\sum_{x}\sum_{y\neq x}[p(x,y)/p(x/y)]$$

$$=\sum_{x}\sum_{y\neq x}q(y)=\sum_{x}[1-q(x)]=\|\boldsymbol{X}\|-1,$$

$$\sum_{x,y}[p(x,y,0)/p(x/y)]$$

$$=\sum_{x}\sum_{y=x}[p(x,y)/p(x/y)]$$

$$=\sum_{x}q(x)=1.$$

(3.28)就是所求证的不等式(3.27)。引理得证。

定理2 （无记忆信道序列的反编码定理）。设$\boldsymbol{C}^n$，$n=1,2,3,\cdots$，是一个由信道$\boldsymbol{C}$生成的无记忆信道序列，则信道容量$R_0=C(\boldsymbol{C})$是$\boldsymbol{C}^n$的一个最大可达速率。这时对任何$R'>R_0$都不能是$\boldsymbol{C}^n$的可达速率。

证明：利用法诺不等式即可证明定理，对此用反证法证。如果有一个$R'>R_0$是$\boldsymbol{C}^n$的可达速率，那么对任何$\varepsilon>0$，总有一

个由(3.5)与(3.6)定义的信源序列$\boldsymbol{S}^n$，其中$M^n=\exp_2[n\cdot R'\cdot(1-\varepsilon)]$，与一个编码序列$(f^n,\ g^n)$，$n=1,\ 2,\ 3,\ \cdots$，使

$$p_e=p_e(f^n,\ g^n)=P_r\{X^n\neq Y^n\}<\varepsilon \tag{3.29}$$

成立，其中$(X^n,\ U^n,\ V^n,\ Y^n)$为由$\boldsymbol{S}^n$，$\boldsymbol{C}^n$，(f^n,g^n)确定是随机变量。这时由法诺(Fano)不等式可得

$$H(X^n/Y^n)\leqslant H(p_e)+p_e\cdot\log(M^n-1)$$
$$\leqslant 1+n\cdot\varepsilon\cdot R'\cdot(1-\varepsilon)$$

成立。这时

$$\begin{aligned}I(X^n;\ Y^n)&=H(X^n)-H(X^n/Y^n)\\&\geqslant\log M^n-1-n\cdot\varepsilon\cdot R'\cdot(1-\varepsilon)\\&=n\cdot R'\cdot(1-\varepsilon)^2-1\end{aligned} \tag{3.30}$$

成立。因为$Y^n=g^n(V^n)$是由V^n确定的随机变量，而X^n与V^n是关于U^n条件独立的随机变量。这时由§3.2引理2可得

$$\begin{aligned}I(X^n;\ Y^n)&\leqslant I(X^n;\ V^n)\leqslant I[(X^n,\ U^n);\ V^n]\\&=I(U^n;\ V^n)\end{aligned} \tag{3.31}$$

成立，其中最后一个等式由柯尔莫各洛夫公式而得。由(3.30)、(3.31)及信道容量的定义可得

$$C(\boldsymbol{C}^n)\geqslant I(U^n;\ V^n)>n\cdot R'\cdot(1-2\varepsilon)-1.$$

如我们选取ε为一个固定的ε_0，使$0<\varepsilon_0<\min[1/8,(R'-R_0)/4]$，那么

$$\begin{aligned}R'\cdot(1-2\varepsilon_0)&\geqslant(R_0+4\varepsilon_0)(1-2\varepsilon_0)=R_0+2\varepsilon_0-8\varepsilon_0^2\\&\geqslant R_0+\varepsilon_0\end{aligned}$$

成立。这时

$$C(\boldsymbol{C}^n)\geqslant n\cdot(R_0+\varepsilon_0)-1.$$

因此当n充分大时必有 $C(\boldsymbol{C}^n)>n\cdot R_0=n\cdot C(\boldsymbol{C})$ 成立，这与§2.2的定理1矛盾。因此$R'>C(\boldsymbol{C})$不能是$\boldsymbol{C}^n$的可达速率。定理得证。

由定理1、定理2的结果可知，无记忆信道的最大可达速率就是它的信道容量．这两个定理反映了无记忆信道在信息传递中的基本特征，因此在许多文献中，把这两个定理称为信道编码的基本定理．信道编码的理论涉及问题很多，例如，有记忆信道编码定理，多用户信道的信道编码定理；信道容量的计算问题；信道编码的误差估计问题及信道编码的工程实现问题等．对信道编码的工程实现问题我们在下节中作一个简介，对多用户信道的编码问题在第五章介绍．而对其余的问题本书就不作详述了，有兴趣的读者可参阅有关文献．

§3.4 纠错码简介

1.编码理论的工程实现问题

在§3.3中，我们给出了信道编码的基本定理，该定理反映了信道在理想情形下的信息传输能力．这个理想情形就是指在保证通信质量（消息传递误差很小）的条件下，传送尽可能多的信息．为实现这个目的，从以上基本定理的证明可知，信道编码运算必需在信号段上进行，这种编码方式在通信工程中称之为分组编码．定理1、定理2告诉我们，这种理想状态的编码是存在的．但另一方面，从工程应用的角度来看，仅停留在存在性问题的讨论上是不够的，工程师们不仅要关心理想情形下的指标，而且更关心的是如何实现这些指标．当分组码的长度n较大时，U^n空间的结构是十分复杂的，§3.3的定理1证明中的编码与译码算法在计算机上都是无法实现的．编码理论的工程实现问题要求编码与译码函数的算法都是易计算的．对此必需借助于代数工具．

为了简单起见，在本节中我们取$\boldsymbol{U}=\boldsymbol{V}=\boldsymbol{F}_2=\{0,1\}$，它们分别代表无、有脉冲信号，且为一个二元域，相应的四则运算可由一定的逻辑电路运算实现。这时n-重信号空间为$\boldsymbol{U}^n=\boldsymbol{V}^n=\boldsymbol{F}_2^n$，相应的信号向量为

$$u^n=(u_1,\ u_2,\ \cdots,\ u_n),\ v^n=(v_1,\ v_2,\ \cdots,\ v_n),$$

其中u_i，$v_i=0$或1，$i=1$，2，…，n。我们以下记$\phi^n=(0,\ 0,\ \cdots,\ 0)$为$n$-维零向量。

2.线性分组码与它的纠错能力

我们称$\boldsymbol{U}^n$的一个子集$\boldsymbol{U}_0^n$为一个分组码或n-维分组码。对任何u，$v\in\boldsymbol{U}$，定义

$$d_H(u,\ v)=\begin{cases}0, & \text{如果}u=v,\\ 1, & \text{如果}u\neq v.\end{cases}$$

而称

$$d_H(u^n,\ v^n)=\sum_{i=1}^{n}d_H(u_i,\ v_i) \tag{4.1}$$

为u^n与v^n的汉明(Hamming)距离，而称

$$d_H(u^n)=\sum_{i=1}^{n}d_H(u_i,\ 0) \tag{4.2}$$

为向量u^n的汉明势。如果$\boldsymbol{U}_0^n$为$\boldsymbol{U}^n$的一个分组码，那么我们称

$$d_H(\boldsymbol{U}_0^n)=\min\{d_H(u^n,\ v^n):\ u^n\neq v^n\in\boldsymbol{U}_0^n\} \tag{4.3}$$

为分组码$\boldsymbol{U}_0^n$的汉明距离。

由以上的定义可知，向量u^n，v^n的汉明距离就是u^n，v^n中不同分量的个数，而向量u^n的汉明势就是u^n中非零分量的个数。

定义1 (1) 称分组码$\boldsymbol{U}_0^n$具有t-纠错能力（或称t-纠错码），如果对发送任何$\boldsymbol{U}_0^n$中的一个向量，在传递误差不超过t个分量时，存在适当的译码方法，能自动纠正误差。

(1) 称分组码$\boldsymbol{U}_0^n$具有s-检错能力(或称s-检错码),如果发送$\boldsymbol{U}_0^n$中的任何一个向量，在传递差错不超过s个分量时，存在适当的译码方法，能自动发现误差。

t-纠检错码与s-检错码的数学描述是这样。如果有一个从$\boldsymbol{V}^n$到$\boldsymbol{U}_0^n$的映射g^n，对任何$v^n \in \boldsymbol{V}^n$，$u^n \in \boldsymbol{U}_0^n$，有如下性质：

(1) 如果$d_H(u^n, v^n) \leqslant t$，那么$g^n(v^n)=u^n$，这时$\boldsymbol{U}_0^n$为$t$-纠错码。

(2) 如果$d_H(u^n, v^n) \leqslant s$，那么$v^n \notin \boldsymbol{U}_0^n$，这时$\boldsymbol{U}_0^n$为$s$-检错码。

以上这种码的自动纠错与检错的功能是信息处理中的重要特征，在通信工程、计算机设计及各种信号的数值处理中有广泛的应用。我们以下讨论这种自动纠错与检错功能的具体实现问题。

定理1 (1) 分组码$\boldsymbol{U}_0^n$是一个t-纠错码的充要条件是

$$t \leqslant \langle [d_H(U_0^n)-1]/2 \rangle, \tag{4.4}$$

其中$\langle a \rangle$表示a的整数部分。

(2) 分组码$\boldsymbol{U}_0^n$是一个s-检错码的充要条件是$s < d_H(\boldsymbol{U}_0^n)$。

证明：如果$\langle [d_H(\boldsymbol{U}_0^n)-1]/2 \rangle < t$成立。此即$d_H(U_0^n) \leqslant 2 \cdot t$。这时必有一对向量$u_1^n$，$u_2^n \in \boldsymbol{U}_0^n$，使$d=d_H(u_1^n, u_2^n) \leqslant 2 \cdot t$成立。不失一般性，我们设向量$u_1^n$与$u_2^n$的前$d$个分量不同，我们作向量$v^n=(v_1, v_2, \cdots, v_n)$为

$$v_i=\begin{cases} u_{1,i}, & \text{如} i=1,2,\cdots,d', \\ u_{2,i}, & \text{如} i=d'+1,d'+2,\cdots,n, \end{cases}$$

其中$d'=\langle (d-1)/2 \rangle +1 \leqslant t$。且有$d_H(u_1^n, v^n)$，$d_H(u_2^n, v^n) \leqslant t$成立。因此$\boldsymbol{U}_0^n$不为$t$-纠错码。反之，如果$\langle [d_H(\boldsymbol{U}_0^n)-1]/2 \rangle \geqslant t$成立。这时必有$d_H(\boldsymbol{U}_0^n) \geqslant 2 \cdot t+1$。如我们记

$$\boldsymbol{U}_0^n=\{u_k^n, k=1,2,\cdots,M^n\},$$

且定义

$$\boldsymbol{O}(u_k^n,t)=\{v^n\in\boldsymbol{U}^n:d_H(u_k^n,v^n)\leqslant t\},\ k=1,2,\cdots M^n,$$

这时$\boldsymbol{O}(u_k^n,\ t)$，$k=1,2,\cdots,M^n$互不相交．如作

$$g^n(v^n)=\begin{cases}u_k^n, & \text{如}v^n\in\boldsymbol{O}(u_k^n,\ t),\ k=1,2,\cdots,M^n,\\ \text{任取}\boldsymbol{U}_0^n\text{中一个向量，否则．}\end{cases}$$

这时$\boldsymbol{U}_0^n$必可纠正t个误差．命题(1)得证．而命题(2)是显然的．定理得证．

3.线性码的构造

一个分组码$\boldsymbol{U}_0^n$被称为$(n,\ m)$-阶线性码，如果$\boldsymbol{U}_0^n$是$\boldsymbol{U}^n$的一个m-维线性子空间，则由二元域上的线代数理论可知，这时必有一个$m\times n$-阶矩阵

$$\boldsymbol{G}=\begin{pmatrix} g_{1,1} & g_{1,2} & \cdots & g_{1,n}\\ g_{2,1} & g_{2,2} & \cdots & g_{2,n}\\ \cdots & \cdots & \cdots & \cdots\\ g_{m,1} & g_{m,2} & \cdots & g_{m,n}\end{pmatrix} \qquad (4.5)$$

其中$g_{i,j}=0$或1，$i=1,\ 2,\ \cdots,\ m$；$j=1,\ 2,\ \cdots,\ n$，且使

$$\boldsymbol{U}_0^n=\{u^n=x^m\cdot\boldsymbol{G}:\ x^m\in\boldsymbol{F}_2^m\} \qquad (4.6)$$

成立，其中$x^n\cdot\boldsymbol{G}$为二元域上的矩阵积．这时称$\boldsymbol{G}$为(n,m)-阶线性码$\boldsymbol{U}_0^n$的生成矩阵．

如果u^n与v^n是$\boldsymbol{F}_2^n$中的两个向量，记$(u^n,\ v^n)=\sum_{i=1}^{n}u_i\cdot v_i$为$u^n$，$v^n$向量的内积．如$(u^n,\ v^n)=0$,则称向量$u^n$与$v^n$相互正交．

如果向量v^n与$\boldsymbol{U}_0^n$中的每个向量正交，则称向量v^n与$\boldsymbol{U}_0^n$正交．如果$\boldsymbol{V}_0^n$由全体与$\boldsymbol{U}_0^n$正交，那么称$\boldsymbol{V}_0^n$与$\boldsymbol{U}_0^n$正交的向量组成，那么称$\boldsymbol{V}_0^n$是$\boldsymbol{U}_0^n$的一个正交空间．

由线性代数理论可知，如果$\boldsymbol{U}_0^n$是一个(n, m)-阶线性码，那么它的正交空间$\boldsymbol{V}_0^n$也是$\boldsymbol{U}^n$中的$(n-m)$-维线性子空间。它的生成矩阵为

$$\boldsymbol{H}=\begin{pmatrix} h_{1,1}, & h_{1,2}, & \cdots, & h_{1,n} \\ h_{2,1}, & h_{2,2}, & \cdots, & h_{2,n} \\ \cdots\cdots & \cdots\cdots & \cdots\cdots & \cdots\cdots \\ h_{k,1}, & h_{k,2}, & \cdots, & h_{k,n} \end{pmatrix} \tag{4.7}$$

其中$k=n-m$。这时称$\boldsymbol{H}$为线性码$\boldsymbol{U}_0^n$的校验矩阵，且有$u^n\in\boldsymbol{U}_0^n$的充要条件是$u^n\cdot\boldsymbol{H}=\phi^n$。

定理2 如果$\boldsymbol{H}$为线性码$\boldsymbol{U}_0^n$的校验矩阵，且$\boldsymbol{H}$的任何t个列线性无关，而有$(t+1)$个列线性相关，那么$d_H(\boldsymbol{U}_0^n)=t+1$，因此$\boldsymbol{U}_0^n$是一个$\langle t/2\rangle$-纠错码。

这个定理由上述线性码与校验矩阵的性质直接推出，对此不再详述。

4.线性循环码

线性循环码是线性码中最重要的一类码，它的编码运算可由移位寄存器实现。因此它的硬件构造十分简单。在本书中，我们只介绍它的一些基本知识。

一个n-维向量$a^n=(a_1, a_2, \cdots, a_n)$，我们定义它的一个循环置换运算为

$$\tau(a^n)=(a_2, a_3, \cdots, a_n, a_1). \tag{4.8}$$

定义2 一个线性码$\boldsymbol{U}_0^n$被称为循环码，如果对任何$a^n\in\boldsymbol{U}_0^n$，那么必有$\tau(a^n)\in\boldsymbol{U}_0^n$。

循环码的构造可由有限域上的多项式环的理想来实现。因为它的描述涉及许多代数方面的知识与术语，本书就不作详述。

重要的循环码有如汉明码、BCH码等。汉明码是一个$n=2^k$

-1，$m=n-k$的(n, m)-阶循环码，它的校验矩阵$\boldsymbol{H}_k$的全体列向量是$\boldsymbol{F}_2^k$的全体非零向量。

例如，在$k=3$时的汉明码是一个（7，4)-阶循环码。相应的校验矩阵为

$$\boldsymbol{H}_3=\begin{pmatrix} 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 \end{pmatrix}$$

由定理 2 可知，汉明码能纠正 1 个误差。它的译码算法如下：

(1) 对任何接收信号v^n，计算$s^m=v^n\cdot\boldsymbol{H}^T$，其中$\boldsymbol{H}^T$为矩阵$\boldsymbol{H}$的转置。

(2) 如果$s^m=\phi^m$，那么作译码函数为$g^n(v^n)=v^n$。

(3) 如果$s^m\neq\phi^m$，那么s^m必为$\boldsymbol{H}_k$的一个列，如记为第i列。那么作译码函数

$$g^n(v^n)=(v_1, \cdots, v_{i-1}, v_i+1, v_{i+1}, \cdots, v_n),$$

这时译码函数g^n必可纠正汉明码的一个误差。

汉明码的推广是BCH码、戈帕(Goppa)码等。

第四章　信息量的推广与应用

§4.1　连续分布的仙农熵与互熵

1.仙农熵的推广问题

从前两章的讨论可以看出，仙农熵在信源与信道编码理论中起着关键性的作用，人们由此可以看出仙农熵的重要意义．但从另一个角度来讲，由于信息科学在近代社会中的重大作用，因此人们对信息度量的范围与应用的期望往往更大些，希望它能在更多的学科分支中发挥作用．因此关于信息量的推广问题一直是许多科学家关心的问题．

关于信息量的推广问题，大体上按以下两个方向进行，即由离散随机变量的仙农熵推广到具有连续分布随机变量的仙农熵，及把仙农熵中的对数函数推广为一般的凸函数．对这两种形式的推广，许多学者作了大量的工作，得到了许多有意义的结果．这些结果虽不像仙农熵那样完整，但是作为信息量的研究发展，还是值得我们学习重视的．

由离散的仙农熵到连续分布的仙农熵的推广问题，在本质上就有困难出现．因为如果把§1.2中的仙农熵的定义式(2.5)作形式的推广，人们发现H(X) 很快就会趋于无穷大．因此，人们把连续分布的信息量的定义基础不是从仙农熵出发，而是从“互熵”的定义出发．因此在本节中，我们首先介绍互熵的定义，而且尽可能避开测度空间的有关描述．因为互熵的一般定义是建立在测度空间基础之上的．

2.离散分布的互熵与它的性质

在§1.2中，我们已说明了仙农熵的本质是随机变量的不肯定性，它由随机变量的概率分布确定．而互熵的概念与仙农熵不同，它是两个概率分布的“差异程度”的度量．对此我们进行数学描述，为了简单起见，我们先讨论离散情形．

设$\boldsymbol{Z}=\{1, 2, \cdots, n\}$是一个有限集合．而

$$P=\{p_1, p_2, \cdots, p_n\}, \quad Q=\{q_1, q_2, \cdots, q_n\} \tag{1.1}$$

是$\boldsymbol{Z}$上的两个概率分布，这时

$$p_i, q_i \geqslant 0, \quad i=1, 2, \cdots, n, \text{ 且} \sum_{i=1}^{n} p_i=\sum_{i=1}^{n} q_i=1.$$

定义1 如果P、Q是$\boldsymbol{Z}$上的两个概率分布，则称

$$H(P \parallel Q)=\sum_{i=1}^{n} p_i \cdot \log(p_i/q_i) \tag{1.2}$$

为P关于Q的互熵．

对互熵(1.2)在有些文献中又称为“散度”．由§1.2的引理3可知，对任何有限集合$\boldsymbol{Z}$及$\boldsymbol{Z}$上的任何两个概率分布P、Q必有

$$H(P \parallel Q) \geqslant 0 \tag{1.3}$$

成立，且等号成立的充要条件是$P=Q$（也就是，对任何 $i=1$，2，…，n，有$p_i=q_i$成立）。不等式（1.3）反映了概率分布Q与P的差异度，这个特征在信息统计理论中有重要作用。

我们取$\boldsymbol{Z}=\boldsymbol{X}\times\boldsymbol{Y}$为两个有限集合的积，而取

$$P=[\boldsymbol{X}\times\boldsymbol{Y},\ p(x,\ y)],\ Q=[\boldsymbol{X}\times\boldsymbol{Y},\ p(x)q(y)],$$

为$\boldsymbol{X}\times\boldsymbol{Y}$上的两个概率分布，其中$p(x)=\sum_x p(x,\ y)$，$q(y)=\sum_y p(x,\ y)$。设$X$、$Y$为两个随机变量，具有联合概率分布为$p(x,\ y)$，那么$X$与$Y$的交互信息就是$P$关于$Q$的互熵，也就是$H(P\parallel Q)=I(X;\ Y)$成立。因此，交互信息是一种特殊的互熵。

3.关于连续分布的互熵

以下我们将离散情形下定义的互熵推广到连续分布的情形。为了简单起见，我们取$\boldsymbol{Z}=[a,\ b]$为一个实数区间，$\boldsymbol{Z}$上的概率分布为

$$F=F(x),\ P=P(x),\ Q=Q(x),\ a\leqslant x\leqslant b. \quad (1.4)$$

我们在此基础上讨论互熵的推广。有关性质可利用测度论工具推广到更一般的连续空间中去，如$\boldsymbol{Z}=\boldsymbol{R}^n$为$n$-维实数空间等情形。

对已给的区间$\boldsymbol{Z}=[a,\ b]$，我们记

$$\boldsymbol{A}=\{a_0,\ a_1,\ \cdots,\ a_m\} \quad (1.5)$$

为$\boldsymbol{Z}$的一组分割点，如有

$$a=a_0<a_1<a_2<\cdots<a_{m-1}<a_m=b \quad (1.5')$$

成立。且记

$$\Delta_1=[a_0,\ a_1],\ \Delta_i=(a_{i-1},\ a_i],\ i=2,\ 3,\ \cdots,\ m, \quad (1.6)$$

其中$[a, b]$，$(a, b]$分别为闭区间与半开半闭区间。这时Δ_i，$i=1, 2, \cdots, m$，互不相交，且有$\boldsymbol{Z}=\bigcup_{i=1}^{m}\Delta_i$成立.以下记

$$\delta(\boldsymbol{A})=\max\{(a_i-a_{i-1}); \ i=1, 2, \cdots, m\}. \quad (1.7)$$

设P、Q为$\boldsymbol{Z}=[a, b]$上的概率分布，这时$P=P(x)$，$Q=Q(x)$为区间$[a, b]$上的单增左连续函数，且

$P(a)=Q(a)=0$，$P(b)=Q(b)=1$。这时记

$$P(\Delta_i)=P(a_i)-P(a_{i-1}), \quad Q(\Delta_i)=Q(a_i)-Q(a_{i-1}). \quad (1.8)$$

以下称

$$H(P\parallel Q; \ \boldsymbol{A})=\sum_{i=1}^{m}P(\Delta_i)\cdot\log[P(\Delta_i)/Q(\Delta_i)] \quad (1.9)$$

分P关于Q在分割$\boldsymbol{A}$上的互熵。因为

$[P(\Delta_i), i=1, 2, \cdots, m]$，$[Q(\Delta_i), i=1, 2, \cdots, m]$

可看作两个离散型的概率分布，因此$H(P\parallel Q; \boldsymbol{A})$是一个离散型分布的互熵。以下记

$$\boldsymbol{B}=\{b_0, b_1, \cdots, b_m\} \quad (1.5'')$$

也为$\boldsymbol{Z}$上的一个分割，对b_j，$j=0, 1, \cdots, m'$，同样有（1.5）的不等式成立。而且同样可定义分布P关于Q在$\boldsymbol{B}$上的互熵。

定义2 如$\boldsymbol{A}$与$\boldsymbol{B}$是$\boldsymbol{Z}$上的两个分割，如果$\boldsymbol{A}$是$\boldsymbol{B}$的一个子集，那么称分割$\boldsymbol{B}$是$\boldsymbol{A}$的一个细分。

引理1 如果$\boldsymbol{B}$是$\boldsymbol{A}$的一个细分，那么必有

$$H(P\parallel Q; \ \boldsymbol{A})\leqslant H(P\parallel Q; \ \boldsymbol{B}) \quad (1.10)$$

成立。

证明：对分割$\boldsymbol{B}$同样可定义

$\Delta_1'=[b_0, b_1]$，$\Delta_j'=(b_{j-1}, b_j]$，$j=2, 3, \cdots m'$，

$$(1.6')$$

如果$\boldsymbol{B}$是$\boldsymbol{A}$的一个细分，那么对每个Δ_i必为若干个Δ'_j分割而成。这时由$H(P\|Q;\ \boldsymbol{A})$与$H(P\|Q;\ \boldsymbol{B})$的定义及§1.2中的不等式（2.8）即可推出不等式（1.7）式成立。引理得证。

定义3 如记$\boldsymbol{D}[a,\ b]$为区间$[a,b]$上的全体有限分割。P、Q为区间$[a,\ b]$上的两个概率分布，那么我们定义

$$H(P\|Q)=\sup\{H(P\|Q;\ \boldsymbol{A});\ \boldsymbol{A}\in\boldsymbol{D}[a,\ b]\} \tag{1.11}$$

为P关于Q的互熵。

定理1 关于互熵有以下性质成立。

（1）对任何$\boldsymbol{Z}=[a,\ b]$上的两个概率分布P、Q总有$H(P\|Q)\geqslant 0$成立，且等号成立的充要条件是$P\equiv Q$，也就是，对任何$x\in[a,\ b]$，$P(x)=Q(x)$。

（2）如记$\boldsymbol{A}_n=\{a_0,\ a_1,\ \cdots,\ a_n\}$，$n=1,\ 2,\ \cdots$，为$[a,\ b]$区间的一列分割，如果$\delta_n=\delta(\boldsymbol{A}_n)\to 0$，则有

$$\lim_{n\to\infty}H(P\|Q;\ \boldsymbol{A}_n)=H(P\|Q). \tag{1.12}$$

（3）如果Q是一个勒贝格测度，也就是$Q(x)=(x-a)/(b-a)$，而$P(x)$是连续分布，具分布密度$p(x)=P'(x)$（$P'(x)$为$P(x)$的导数），这时

$$H(P\|Q)=\int_a^b p(x)\log[p(x)/(b-a)]dx$$

$$=\log(b-a)+\int_a^b p(x)\log p(x)dx. \tag{1.13}$$

如果P、Q具有分布密度$p(x)=P'(x)$与$q(x)=Q'(x)$，那么这时有

$$H(P\|Q)=\int_a^b f(x)\log[f(x)/g(x)]dx. \tag{1.13'}$$

定理1的证明要用到测度论的许多定义与性质，对此我们不作详述. 关于互熵的定义可推广到一般概率空间的情形，因此对n-维实数空间，相应的（1.13′）式为

$$H(P \| Q)=\int_{R^n} p(x^n)\log[p(x^n)/q(x^n)]dx^n, \quad (1.13'')$$

其中$x^n=(x_1, \cdots, x_n)$为n-维实数向量，$dx^n=dx_1\times dx_2\times\cdots\times dx_n$，而$p(x^n)$，$q(x^n)$分别为$P$、$Q$的分布密度。

4.连续分布的仙农熵

作为离散随机变量仙农熵的形式推广，人们把

$$H(x)\doteq-\int_R p(x)\log p(x)dx^n \quad (1.14)$$

称为连续随机变量的仙农熵，其中$p(x)$为随机变量X的分布密度。关于连续随机变量仙农熵的来源与意义不如离散随机变量仙农熵那样明确，它的值还可能是一个负值. 但是（1.14）的仙农熵还是一种很有用的信息量. 它的应用有如以下情形：

（1）关于信源编码理论中有关信号体积的性质可推广到连续随机变量的情形. 如记X_1，X_2，$\cdots$，X_n为一组独立同分布的随机变量，每个X_i具有分布密度$p(x)$，$x\in \mathbf{R}$，这时$X^n=(X_1, X_2, \cdots, X_n)$的概率分布取值集中在

$$p(x^n)=p(x_1)p(x_2)\cdots p(x_n)\backsim \exp_2[-n\cdot H(X)] \quad (1.15)$$

的范围内.

（2）利用连续随机变量的仙农熵同样可产生条件熵与交互信息. 如记随机变量(X, Y)具有联合分布$p(x, y)$与相应的边际分布$p(x)$与$q(y)$，那么它们的条件熵与交互信息可定义为

$$H(X/Y)=-\int_{R^2}p(x,y)\log[p(x,y)/q(y)]dxdy, \tag{1.16}$$

$$H(Y/X)=-\int_{R^2}p(x,y)\log[p(x,y)/p(x)]dxdy, \tag{1.16'}$$

$$I(Y;X)=\int_{R^2}p(x,y)\log[p(x,y)/(p(x)q(y))]dxdy. \tag{1.17}$$

这时，交互信息是一种特殊的互熵，因此它的意义是明确的。

5.几种典型分布的仙农熵

在概率统计中，最常见的几种分布有如均匀分布、指数分布、正态分布与一般指数型分布，它们的密度函数为：

(1) 均匀分布：$p_1(x)=\begin{cases}1/(b-a), & 如果x\in[a,b],\\0, & 否则.\end{cases}$ (1.18)

(2) 指数分布：$p_2(x)=\begin{cases}\lambda\exp(-\lambda x), & 如果x>0,\\0, & 否则.\end{cases}$ (1.19)

(3) 正态分布：$p_3(x)=\frac{1}{\sqrt{2\pi}\cdot\sigma}\cdot$

$\exp\{-[(x-\mu)^2/\sigma^2]\}.$ (1.20)

(4) 一般指数型分布：

$p_4(x)=C\cdot\exp(\tau_1x+\tau_2x^2+\cdots+\tau_kx^k),$ (1.21)

其中 C为归一化系数，使$\int p_4(x)dx=1$。

众所周知，正态分布是一种特殊的指数型分布。

(5) 多维正态分布：

$$p(x^n)=\left[\frac{\|a_{i,j}\|}{2\pi}\right]^{1/2}\exp\left\{-\frac{1}{2}\sum_{i,j=1}^{n}a_{i,j}(x_i-\mu_i)\cdot(x_j-\mu_j)\right\}, \quad (1.22)$$

其中$\|a_{i,j}\|$为矩阵$(a_{i,j})$的行列式，矩阵$(a_{i,j})$是相关矩阵$(\sigma_{i,j})$的逆矩阵，而

$$\mu_i=\int_{R^n}x_i p(x^n)dx^n,$$

$$\sigma_{i,j}=\int_{R^n}(x_i-\mu_i)(x_j-\mu_j)p(x^n)dx^n. \quad (1.23)$$

对这五种概率分布的仙农熵分别为：

$$H(X_1)=\log_2(b-a); \quad (1.24)$$

$$H(X_2)=\log_2(e/\lambda); \quad (1.25)$$

$$H(X_3)=(1/2)\log_2(2\pi e\cdot\sigma^2); \quad (1.26)$$

$$H(X_4)=\log_2 C+\left(\sum_{i=1}^{k}\tau_i r_i\right)\cdot\log_2 e, \quad (1.27)$$

其中 $r_i=\int x^i p(x)dx$为随机变量的i阶矩，而

$$H(X^n)=(1/2)\log_2[(2\pi e)^n\cdot\|a_{i,j}\|]. \quad (1.28)$$

在(1.24)—(1.27)式中，X_1-X_4分别为具有均匀分布、指数分布、正态分布与一般指数型分布的随机变量，而(1.28)式中的X^n为n-元正态分布的随机变量。

§4.2 最大熵与最小互熵原理

1.最大熵与最小互熵原理

最大熵与最小互熵原理是信息量的两个重要性质，它们深刻地反映了概率分布的信息特征，在近代统计理论中有重要应用，而且它们的作用还在增加。在本节中，我们介绍这些原理。

在介绍最大熵与最小互熵原理之前，我们必须先介绍“先验信息”这个概念．“先验信息”这个名词在统计理论中广为引用，它表示对概率分布（在统计中称为样本总体）的某些已知约束条件．这些条件确定了概率分布的变化范围。因此，本书中所指的先验信息就是概率分布函数的约束条件。常用的约束条件有两种类型，即有分布取值范围的约束与函数矩约束。例如：

约束条件1：分布密度函数$p(x)$在$[a,b]$区间之外为零；

约束条件2：分布密度函数$p(x)$在$x<0$处为零；

约束条件3：设$r_i(i=1,\ 2,\ \cdots,\ k,)$为固定常数，分布密度$p(x)$使关系式

$$\int_R x^i\cdot p(x)dx=r_i,\qquad i=1,2,\cdots,k, \tag{2.1}$$

成立。

约束条件1，2为分布取值范围约束，而(2.1)是矩约束，我们称之为k-阶矩约束．由这些约束条件确定了分布密度函数的变化范围．我们记$\boldsymbol{P}$为$p(x)$的可能变化范围．因此$\boldsymbol{P}$就是先验信息．记$\boldsymbol{P}$中的元为概率分布F，P，Q等，而相应的分布密

度分别为$f(x)$，$p(x)$，$q(x)$。

最大熵分布就是指在先验信息$\boldsymbol{P}$中的一个概率分布P_0，使它的仙农熵在$\boldsymbol{P}$中为最大，这时

$$H(P_0)=\max\{H(P)\colon\ P\in\boldsymbol{P}\}. \tag{2.2}$$

最小熵分布就是指在先验信息$\boldsymbol{P}$中的一个概率分布P_0，使它与Q分布的互熵为最小，这时

$$H(P_0\parallel Q)=\min\{H(P\parallel Q)\colon\ P\in\boldsymbol{P}\}. \tag{2.3}$$

以下记

$\boldsymbol{P}_1$为使约束条件(1)成立的全体连续分布P；

$\boldsymbol{P}_2$为使约束条件(2)与1-阶矩约束成立的全体连续分布P；

$\boldsymbol{P}_3$为使2-阶矩约束条件成立的全体连续分布P；

$\boldsymbol{P}_4$为使一般k-阶矩约束条件成立的全体连续分布P；

$\boldsymbol{P}_5$为使2-阶相关矩约束条件成立的全体连续分布P。2-阶相关矩约束条件是使(1.23)成立的全体$p(x^n)$，这时$\mu_i,\sigma_{i,j},i,j=1,2,\cdots,n$。

最大熵原理 如记P_i为$\boldsymbol{P}_i$中的最大熵分布，$i=1,2,3,4,5$。这时P_1—P_5的分布密度$p_1(x)$—$p_4(x)$与$p_5(x^n)$分别为由(1.18)—(1.22)式定义的均匀分布，指数分布，正态分布，指数型分布与多维正态分布，且它们的参数由约束条件决定。

最小互熵原理 如记P_i为$\boldsymbol{P}_i$中与分布Q的互熵为最小的分布，且记它们的分布密度分别为$p_i(x)$，$i=1,2,3,4$与$p_5(x^n)$。这时P_i与Q的分布密度的比$p_i(x)/q(x)$，$i=1,2,3,4$与$p_5(x^n)/q(x^n)$分别为由(1.18)—(1.22)式定义的函数，且它们的参数由约束条件决定。

以上原理可在许多信息论的书中找到，对此我们不再详述。

2.高斯(Gaussian) 信道的信道容量

在第三章中，我们给出了离散信道的信道容量及它的编码定理，我们现在考虑连续状态的信道容量问题. 这时信道$\boldsymbol{C}=[\boldsymbol{U},\ p(v/u),\ \boldsymbol{V}]$，其中$\boldsymbol{U}=\boldsymbol{V}=\boldsymbol{R}$为实数空间，而 $p(v/u)$ 条件概率分布密度. 以下记U，V，Z为信道的输入、输出与噪声随机变量，它们都是在实数空间中取值.

定义1 对上述信道$\boldsymbol{C}$的输入、输出与噪声随机变量U,V,Z，如果U与Z是相互独立的随机变量，且$V=U+Z$，则称信道$\boldsymbol{C}$为加性信道.

如信道$\boldsymbol{C}$是加性信道，且噪声Z是一个高斯随机变量，这时称信道$\boldsymbol{C}$是一个加性高斯信道.

如信道$\boldsymbol{C}$是一个加性高斯信道，那么

$$p(v/u)=(2\pi\cdot N^2)^{-1/2}\exp\{-(v-u)^2/(2\cdot N^2)\},\tag{2.4}$$

其中N^2为噪声功率.

在连续信道$\boldsymbol{C}$中，输入信号随机变量U还未给定，在通信工程中，U 的取值一般为某种物理量，如电流强度、电压等，且对它们的功率要有一定的限制，如

$$\boldsymbol{E}\{U^2\}=\int_R u^2p(u)du\leqslant\sigma^2,\tag{2.5}$$

这时要求输入信号的平均功率不超过σ^2，我们称σ^2为输入信号的最高平均功率.

对一个加性高斯信道$\boldsymbol{C}$，如果输入信号随机变量U的概率分布密度$p(u)$给定，且满足功率约束条件(2.5)，那么输出信号随机变量V的平均功率为

$$\boldsymbol{E}\{V^2\}=\boldsymbol{E}\{(U+Z)^2\}=\boldsymbol{E}\{U^2\}+\boldsymbol{E}\{Z^2\}$$

$$\leqslant \sigma^2 + N^2, \tag{2.6}$$

其中σ^2，N^2分别为输入信号与噪声的平均功率。这时U、V的交互信息为

$$I(U;V) = H(V) - H(V/U) \leqslant (1/2)\log(1 + \sigma^2/N^2). \tag{2.7}$$

(2.7)中的不等式由(2.5)与(2.6)式可得，因为由(2.5)式得

$$H(V/U) = \int_R \int_R p(u,v)\log p(v/u)dvdu$$

$$= (1/2)\log(2\pi e \cdot N^2), \tag{2.8}$$

由(2.6)可得

$$H(V) \leqslant (1/2)\log[2\pi e \cdot (N^2 + \sigma^2)]. \tag{2.9}$$

将(2.8)与(2.9)代入(2.7)的不等式左边就得(2.7)式。

如我们取信道输入随机变量U为高斯随机变量，具正态分布$N(0,\sigma^2)$，那么输出随机变量V也是一个高斯随机变量，具有正态分布$N(0,N^2+\sigma^2)$，因此U、V的交互信息使(2.7)的等号成立。我们称

$$C = (1/2)\log(1 + \sigma^2/N^2) \tag{2.10}$$

为加性高斯信道的信道容量，而σ^2/N^2为信噪比。由此可知，对决定加性高斯信道的信道容量大小的关键是信噪比的大小。

对连续状态下的信道，我们同样可给出它的编码问题。这时，要求编码函数

$$u^n = (u_1, u_2, \cdots, u_n) = f^n(x^n) \tag{2.11}$$

满足功率约束条件：

$$\sum_{i=1}^{n} u_i^2 \leqslant n \cdot \sigma^2 (1 + \varepsilon), \tag{2.12}$$

其中ε为任意小的正数。与§3.3的定义相同，我们同样可给出

连续状态信道的可达速率，这时要求编码函数$f^n(x^n)$满足约束条件(2.12)。

与§3.2的定义相同，我们可同样定义无记忆的连续状态信道。对无记忆加性高斯信道同样有它的编码定理。这时(2.10)的信道容量C是无记忆加性高斯信道的最大可达速率。对此命题的证明与§3.3的正、反编码定理相似，我们不再详述。

§4.3 广义熵的定义与性质

在§4.1的开始，我们就已指出，信息量的推广有着十分重要的意义，它的推广途径沿着把离散仙农熵向连续分布与一般的凸函数的两个方向进行。在§4.1，§4.2中，我们给出了关于连续分布熵的几种定义的引进与性质。在本节中，我们讨论把仙农熵中的对数函数向一般凸函数的推广。以下记$\boldsymbol{X}=\{1, 2, \cdots, n\}$，而

$$\boldsymbol{P}=\{p^n=(p_1, \cdots, p_n):\ p_i\geqslant 0,\ \sum_{i=1}^{n} p_i=1\}. \quad (3.1)$$

这时$\boldsymbol{P}$为一个凸集，也就是对任何p_1^n，$p_2^n\in\boldsymbol{P}$，总有$0.5(p_1^n+p_2^n)\in\boldsymbol{P}$。

定义1 我们称$H(p^n)$为一个广义熵(有的文献中称为多样性指标)，如果$H_G(p^n)$的定义域为$\boldsymbol{P}$，且满足以下条件：

(1) 对任何$p^n\in\boldsymbol{P}$，$H_G(p^n)\geqslant 0$；

(2) 对任何p^n向量的置换T，总有$H_G[T(p^n)]=H_G(p^n)$，其中

$$T(p_1, p_2, \cdots, p_n)=(p_{r(1)}, p_{r(2)}, \cdots, p_{r(n)})$$

对任何$p^n\in\boldsymbol{P}$成立，其中$T(i)$为$\boldsymbol{N}=\{1, 2, \cdots, n\}$上的置换

运算.

(3) $H_G(p^n)$为 $\boldsymbol{P}$ 上的一个严格凸函数. 也就是, 对任何 p_1^n, $p_2^n \in \boldsymbol{P}$, 总有

$$H_G[0.5(p_1^n + p_2^n)] \geq 0.5[H_G(p_1^n) + H_G(p_2^n)] \quad (3.2)$$

成立, 且等号成立的充要条件是$p_1^n = p_2^n$成立.

重要的广义熵有如:

(1) Φ—熵

如果$\phi(p)$ 是一个[0, 1]区间上的凸函数, 对任何 p, $q \in$ [0, 1], 总有

$$\phi[0.5(p+q)] \geq 0.5[\phi(p) + \phi(q)]. \quad (3.3)$$

这时称

$$H_G(p^n) = \sum_{i=1}^{n} \phi(p_i). \quad (3.4)$$

重要的ϕ-熵有:

仙农熵, 当$\phi(p) = -p \cdot \log p$时;

桑普松(Simpson)指数, 当$\phi(p) = p \cdot (1-p)$时, 这时

$$H_G(p^n) = \sum_{i=1}^{n} p_i \cdot (1-p_i) = 1 - \sum_{i=1}^{n} p_i^2. \quad (3.5)$$

桑普松指数在生物学中有重要的应用.

(2) 拟Φ-熵

如果$\phi(p)$是一个[0, 1]区间上的凸函数, ψ是一个凸函数, 这时称

$$H_G(p^n) = \psi\left[\sum_{i=1}^{n} \phi(p_i)\right]. \quad (3.6)$$

重要的拟ϕ-熵有α-阶熵 (或莱尼 (Renyi) 熵). 这时取

$$\psi(u) = [1/(\alpha - 1)] \cdot \log(u), \quad \phi(p) = p^{1-\alpha}, \quad (3.7)$$

因此有

$$H(p^n;\alpha)=[1/(\alpha-1)]\log\Big[\sum_{i=1}^{n}(p_i)^{1-\alpha}\Big]. \qquad (3.8)$$

引理1 仙农熵可看作一种特殊的α-阶熵，也就是在$\alpha\to1$时，α-阶熵变为仙农熵。这时有

$$\lim_{\alpha\to1}H(p^n;\ \alpha)=H(p^n),$$

其中$H(p^n)$为仙农熵。

对广义熵我们同样可定义它们的条件熵与交互信息，如

$$H_G(X/Y)=H_G(X,\ Y)-H_G(Y), \qquad (3.9)$$

$$H_G(Y/X)=H_G(X,\ Y)-H_G(X), \qquad (3.9')$$

$$I_G(X;\ Y)=H_G(X)+H_G(Y)-H_G(X,Y). \qquad (3.10)$$

仙农熵的许多性质可推广到广义熵的情形，对此我们不一一列举了。

第五章 编码理论的发展与应用

§5.1 率失真函数与数据压缩编码定理

1.数据压缩编码问题

在第二章中我们已给出了无噪声时的信源编码定理，给出了两种类型的编码问题，对这两种类型的编码问题，它们的处理方法不同，但是它们的共同目的都是要用最少的信号把信源的大部分信息再现出来．但第二章的信源编码理论只考虑信源的概率分布因素，因此它的应用范围受到限制．为了以更经济的手段把信源的信息再现出来，我们还必需作更进一步的讨论分析．在一个实际的信息处理问题中，还有以下因素可供利用．即有

(1) 对不同的消息，它们之间存在一定的度量关系．以中文字母为例，有些反意字，如“上、下”，“来、去”，“收、支”等字含意差别极大，如它们在通信中出错就会有很大的损失．而有些字，如“和、与、同”等，它们的差别很小，有了错也

不会有太大的损失。对这种不同的度量关系我们可用一个度量函数$d(x, y)$，$x, y \in \boldsymbol{X}$来表示，$d(x, y)$ 代表消息字母x与y之间的差异程度。这时，一个信源可用

$$\boldsymbol{S}=[\boldsymbol{X}, p(x), d(x, y), \boldsymbol{Y}] \tag{1.1}$$

来表示，其中$d(x, y)$ 为用y来复制x时的损失度量。

（2）在文字、语音与图象处理问题中，对实际消息的再现可允许一定的误差存在，例如，在电视图象中，只要有足够多的线素就可保持很好的清晰度，为人们的视觉所接受，同样，在电影的连续图象中，每秒钟只要有足够多的变化画面，就可保持动作的连贯性，为人们的视觉所接受。又如，在卫星照片中，对气象照片与侦察照片的清晰度的要求有很大的差别，因此在传递气象卫星的照片时，数据可作更多的压缩。这种可允许的误差记为D，对不同的信息处理可有不同的误差要求。

数据压缩理论就是在信源复制误差不超过已给的允许度D的要求下，以最少的数据来再现原信源的信息。为了实现这个目的，信源编码理论仍在序列模型下讨论，记

$$\boldsymbol{S}^n=[\boldsymbol{X}^n, p(x^n), d(x^n, y^n), \boldsymbol{Y}^n], n=1, 2, \cdots, \tag{1.2}$$

为一个有度量函数的信源序列。

定义1 我们称常数R是信源序列(1.2)的一个D–可达速率，如果存在一列编码序列（f^n, g^n），$n=1, 2, 3, \cdots$，对任何$\varepsilon>0$，只要n充分大，满足以下条件：

（1）f^n，g^n分别为映射

$$f^n: \boldsymbol{X}^n \to \boldsymbol{M}^n, \quad g^n: \boldsymbol{M}^n \to \boldsymbol{Y}^n, \tag{1.3}$$

其中$\boldsymbol{M}^n=\{1, 2, \cdots, M^n\}$，$M^n$ 为任何$\leqslant \exp_2[n \cdot R \cdot (1+\varepsilon)]$的数。

（2）复制信源与原信源的畸变不超过D，也就是有

$$P_r\{d(X^n, Y^n)>D\cdot(1+\varepsilon)\}<\varepsilon. \tag{1.4}$$

成立，其中X^n与Y^n的联合概率分布由信源$\boldsymbol{S}^n$与编码（f^n，g^n）给定，这时

$$p(x^n,y^n)=P_r\{(X^n,Y^n)=(x^n, y^n)\}$$
$$=\begin{cases}p(x^n), & \text{如果} y^n=g^n[f^n(x^n)],\\ 0, & \text{否则.}\end{cases} \tag{1.5}$$

由定义1可知，信源序列的D-可达速率R就是信源序列在具有允许误差D的范围内，信源数据可压缩的比特数大体为$n\cdot R$。信源编码问题就是要找出最小的D-可达速率。

2.率失真函数与信源编码定理

为求信源序列的最小D-可达速率，在下文中，我们讨论$\boldsymbol{S}^n$，$n=1$，2，…，为一个无记忆的信源序列。这时

$$p(x^n)=\prod_{i=1}^{n}p(x_i),d(x^n,y^n)=(1/n)\cdot\sum_{i=1}^{n}d(x_i,y_i), \tag{1.6}$$

对任何$x^n=(x_1, \cdots, x_n)\in\boldsymbol{X}^n$，$y^n=(y_1, \cdots, y_n)\in\boldsymbol{Y}^n$成立。如果(1.5)中的$p(x)$与$d(x, y)$由(1.1)给定，这时我们称(1.2)的信源序列为由信源(1.1)生成的无记忆信源序列。

对（1.1）的信源$\boldsymbol{S}$，如果记$Q=[q(y/x)]$为一个从$\boldsymbol{X}\to\boldsymbol{Y}$的转移概率矩阵，那么由$p(x)$，$Q$确定联合概率分布为$p(x,y)=p(x)q(y/x)$。我们记

$$\boldsymbol{Q}(D)=\{Q=[q(y/x)]: \sum_{x,y}d(x,y)p(x)q(y/x)$$
$$\leqslant D\}, \tag{1.7}$$

而相应的交互信息为

$$I(X;Y)=\sum_{x,y}p(x, y)\log[q(y/x)/q(y)], \tag{1.8}$$

其中X，Y为由联合分$p(x, y)$确定的随机变量.

定义2 对（1.1）的信源$\mathbf{S}$，我们称

$$R(D)=\inf\{I(X;\ Y):\ Q\in\mathbf{Q}(D)\} \tag{1.9}$$

为信源$\mathbf{S}$的率失真函数（或为（R, D)-函数）.

定理1 如果$\mathbf{S}^n$，$n=1, 2, \cdots$，为由$\mathbf{S}$生成的无记忆信源序列，那么$\mathbf{S}^n$的最小D-可达速率为率失真函数$R(D)$.

由定理1可知，在一定的允许误差条件下，无记忆信源的最小数据压缩率是可以计算的，而且它的值就是率失真函数$R(D)$。这个结论在信息处理理论中十分重要，它是数据压缩问题的理论基础，在语音、图象等问题中广为应用. 由于篇幅所限，本定理的证明从略.

3.率失真函数的计算公式

与信道容量的计算相同，率失真函数的计算也可用多元函数求极值的方法计算. 这时作拉格朗日函数为

$$\begin{aligned}L[q(y/x)]&=I(X;\ Y)\\&+\sum_x\mu(x)\sum_y q(y/x)+\lambda\sum_{x,\ y}d(x,y)p(x)q(y/x),\end{aligned} \tag{1.10}$$

其中$I(X;\ Y)$由(1.8)定义，它是$q(y/x)$的函数.为求$R(D)$，只要解方程式

$$\partial L/\partial q(y/x)=0,\ x\in\mathbf{X},\ y\in\mathbf{Y}, \tag{1.11}$$

就可. 对几种特殊分布的率失真函数有以下公式.

（1）二进信源$\mathbf{S}$为$\mathbf{X}=\mathbf{Y}=\{0,\ 1\}$，且$p(1)=p$，$p(0)=1-p$，而$d(x,\ y)$为Hamming距离$d_H(x,\ y)$。这时率失真函数为

$$R(D)=H(p)-H(D),\ 当0\leqslant D\leqslant p\leqslant 0.5时, \tag{1.12}$$

如果$D\geqslant p$，则$R(D)=0$。

（2）具有绝对误差的双边指数分布信源.这时 $\boldsymbol{X}=\boldsymbol{Y}=\boldsymbol{R}$ 为全体实数. 而

$$p(x)=(\alpha/2)\exp(-\alpha|x|), \quad (1.13)$$

$d(x, y)=|x-y|$。这时

$$R(D)=-\log(\alpha D), \quad 0\leqslant D\leqslant 1/\alpha. \quad (1.14)$$

（3）具有均方误差的正态分布信源.这时 $\boldsymbol{X}=\boldsymbol{Y}=\boldsymbol{R}$为全体实数. 而

$$p(x)=(2\pi\sigma^2)^{-1/2}\exp[-(x-\mu)^2/(2\sigma^2)] \quad (1.15)$$

$d(x, y)=(x-y)^2$。这时

$$R(D)=(1/2)\cdot\log(\sigma^2/D), \quad 0\leqslant D\leqslant\sigma^2. \quad (1.16)$$

对其它类型信源的率失真函数还有许多，对此就不一一列举了.

§5.2 多用户通信网络概论

1.多用户通信网络编码的发展简史

在§2.1中，我们给出了通信系通的基本要素与框图，它的基本特点是信源、信道及它们的输入、输出都是单一的.因此这种通信模型比较简单.由于近代通信技术的发展，这种简单模型已不能概括近代通信的技术模型.因此在70年代出现了大量网络式的通信模型.在国际上称之为多用户信息论.我们现在概述它的发展历史与模型特征.

最早的多用互通信模型是1963年仙农提出的二元通信模型，该模型的框图如图5.1所示.它的结构要点有：

（1）通信双方可看作甲、乙两地，每方都具有收、发信号的设备。

（2）通信的甲、乙双方分别在同一地点对收、发信号的情形可随时通告。

由这个二元信道模型我们可进行以下几点特征分析。

（1）通信系统具有多个信源、信道且有多个输入、输出。

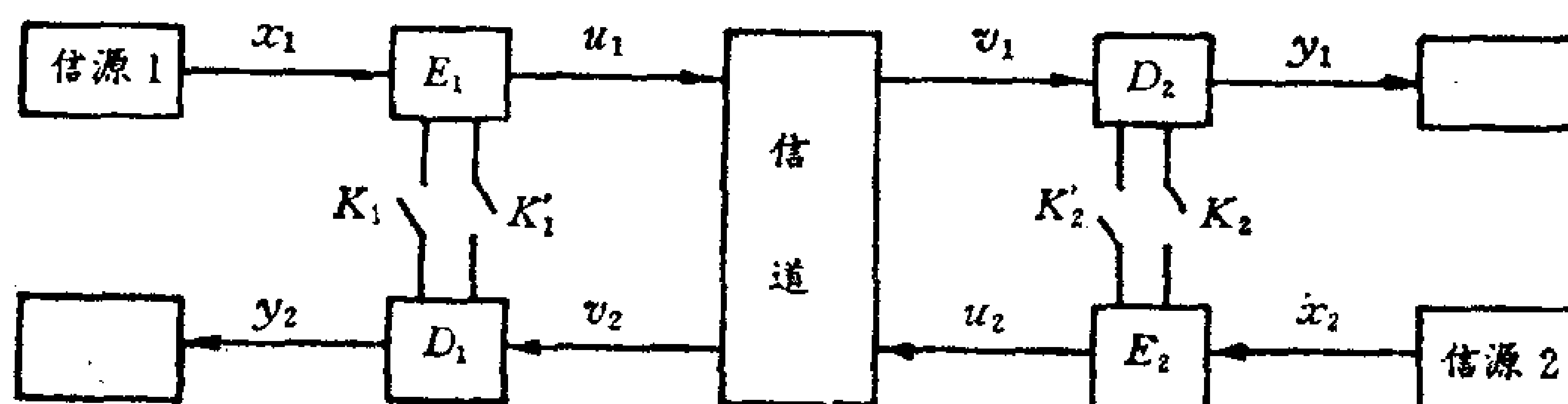


图5.1 二元信道的通信模型

（2）通信过程中信息的传递有多种手段，在图5.1中，除了信道传递信息之外，还有在同一地点内传递的信息。这种信息人们称之为“边信息”与“反馈信息”。这些特点是多用户信息论的基本特征，并由此引出许多其他的模型。

自仙农1963年提出上述二元信道以后，多用户信息论并未立即得到发展，自1963年到70年代初，虽有论文对二元信道进行研究，但多用户信息论的研究进展不大。自70年代初，由于卫星通信技术的发展，提出了一系列多用户通信的网络模型，对它们的编码理论的讨论形成一个高潮，成为这一时期信息论发展的主题之一。

2.多用户信息论的主要模型

多用户信息论的主模型有三大类，即“多重信源的编码模型”、“多路信道的编码模型”及“信源、信道混合编码模型”。

我们现在概述它们的特征。

（1）多重信源的编码模型

多重信源的编码问题是指有多个信源，这些信源具有一定的相关性，它们按一定的编码方式进行信息处理，多重信源编码的基本问题是如何降低它们的可达速率。常见的多重信源模型有如图5.2—图5.7表示的“SWC(Slepian–Wolf–Cover)–模型”；“WAK(Wyner–Ahlswede–Körner)–模型”；“W(Wyner)–

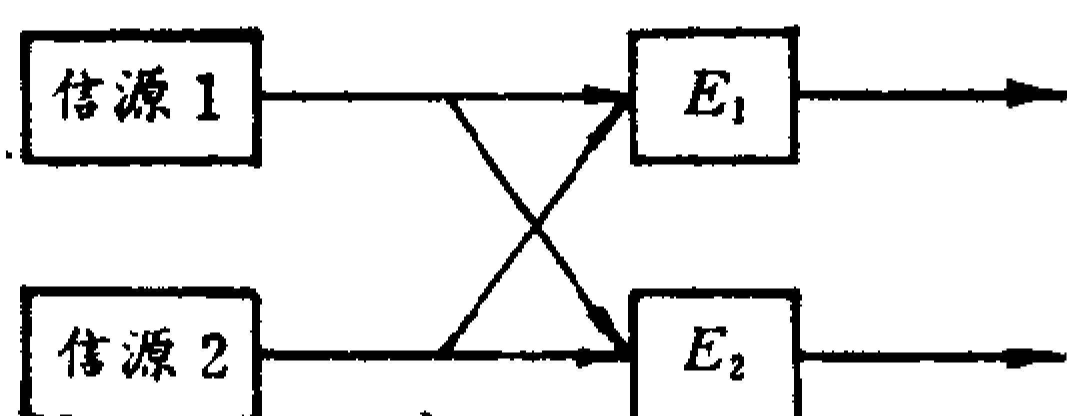


图5.2 SWC—模型

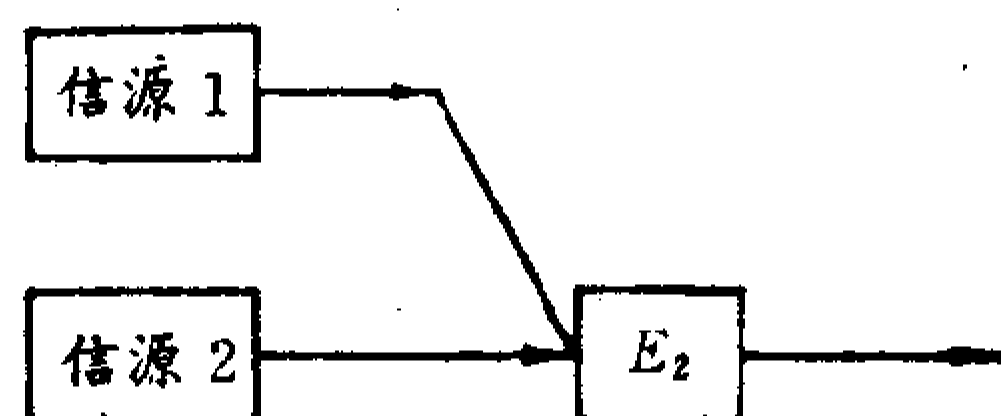


图5.3 WAK—模型

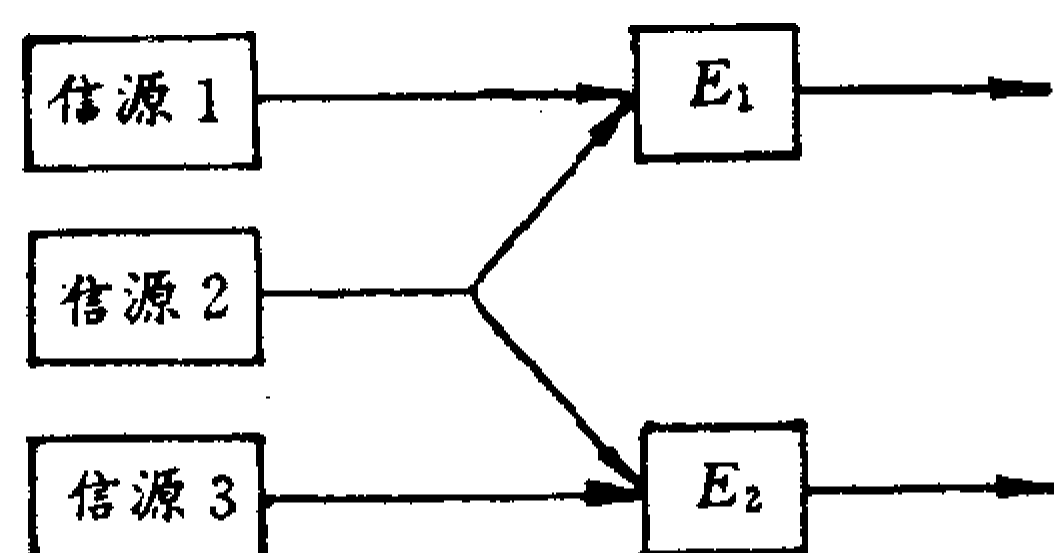


图5.4 W—模型

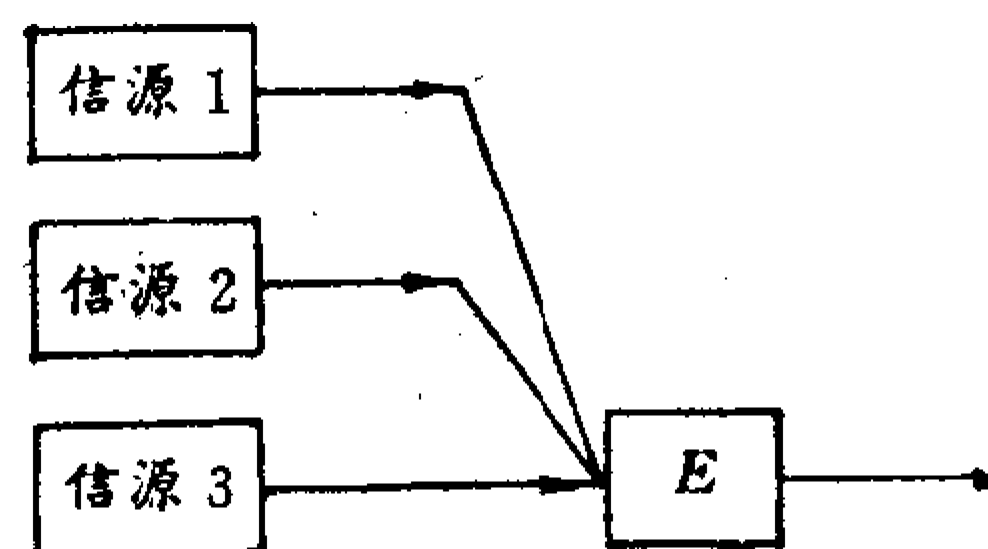


图5.5 S—模型

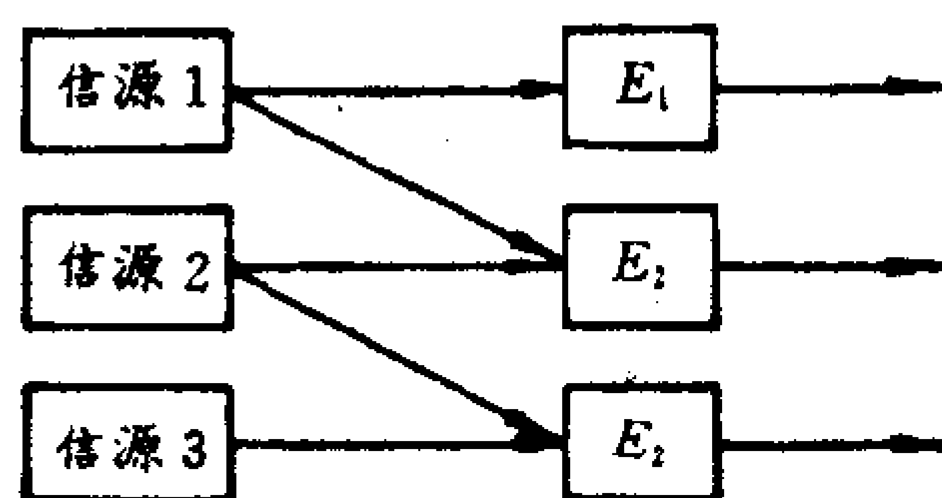


图5.6 KM—模型

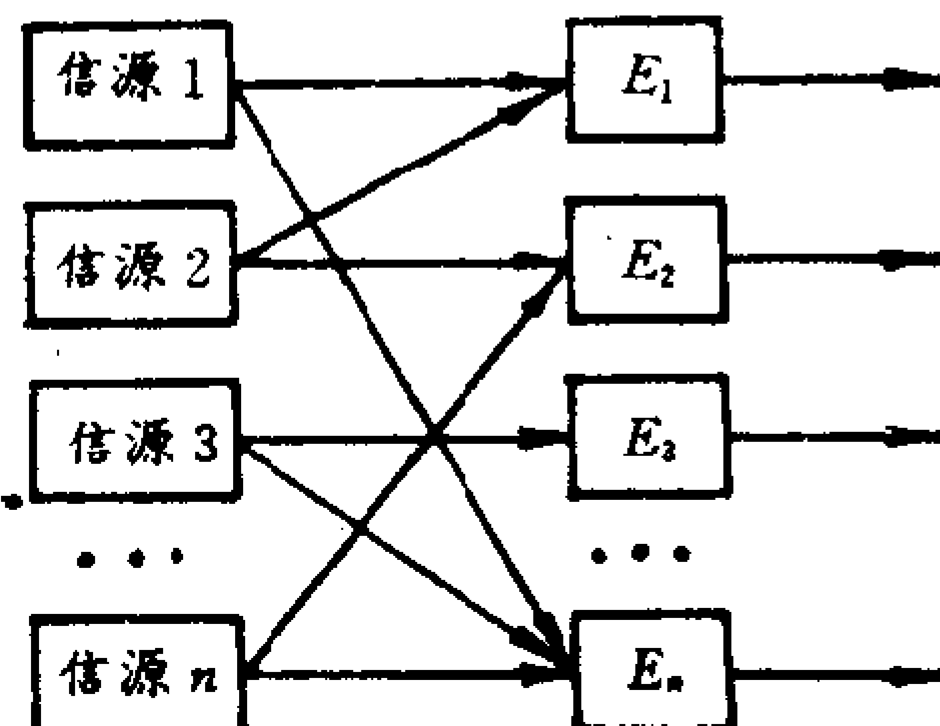


图5.7 HK—模型

模型”；“S (Sgarro)-模型”；“KM(Körner-Marton)-模型”；“HK(Han-Kobayashi)-模型”等。

(2) 多路通信信道

多路通信信道一般指信道有多重输入、输出信号。典型的多路通信信道有如“多址信道”，“广播信道”等，它们的构造为图5.8—图5.11所示。

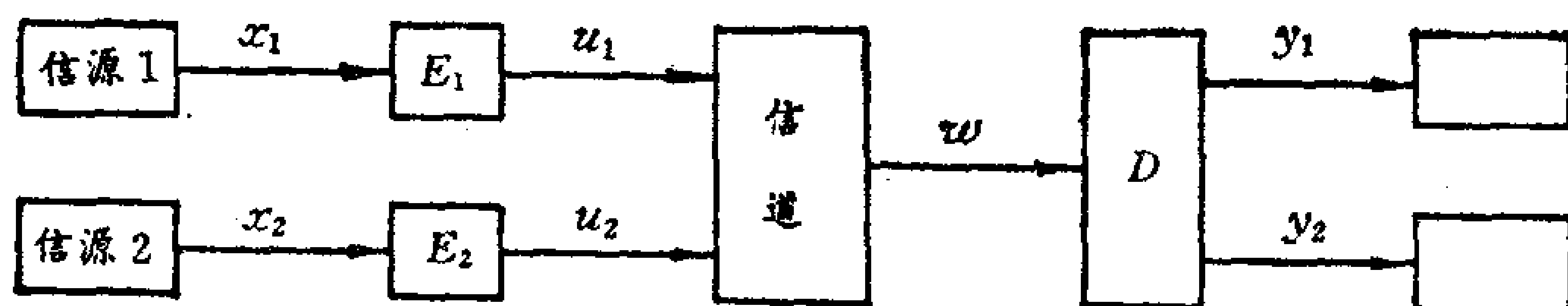


图5.8 二址信道

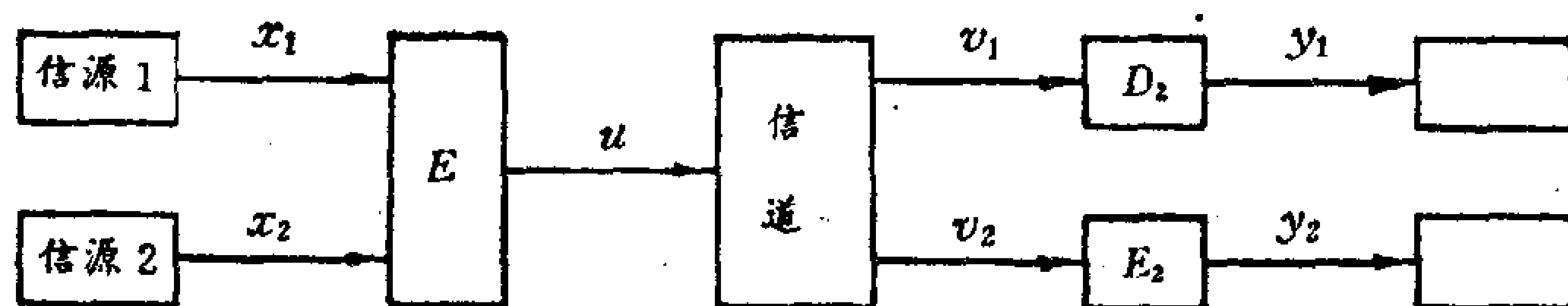


图5.9 标准广播信道

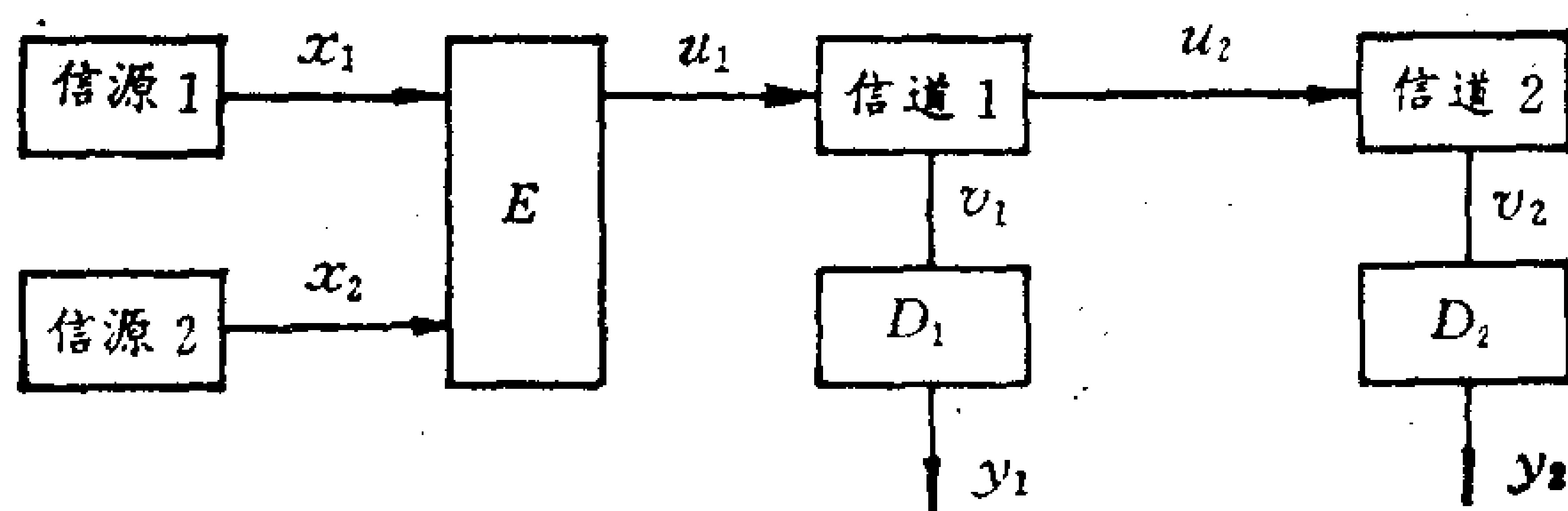


图5.10 降阶广播信道

(3) 混合多用户通信系统

如把上述多重信源与多路信道结合及与率失真理论结合就可产生各种混合多用户通信系统。它的典型模型有如：“有边信息的通信系统”，“有反馈信号的通信系统”，“相关信源与多路

信道的通信系统"，"有允许误差的多用户通信系统"等。其中有边信息与有反馈的通信模型如图5.11、图5.12所示。

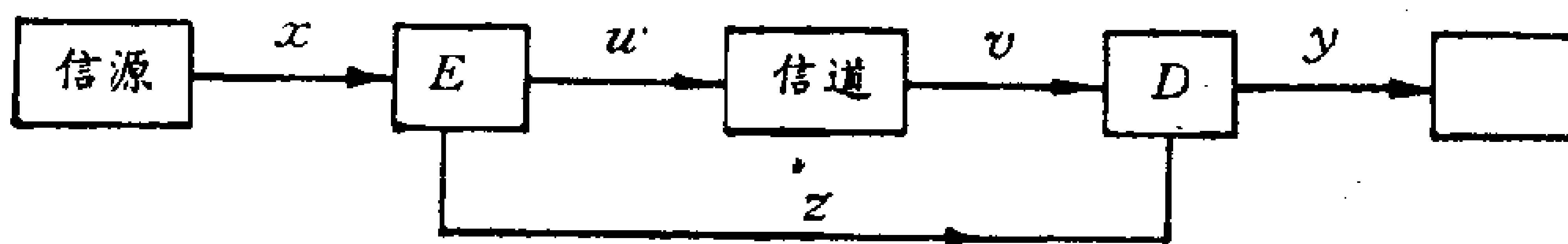


图5.11 有边信息的通信系统

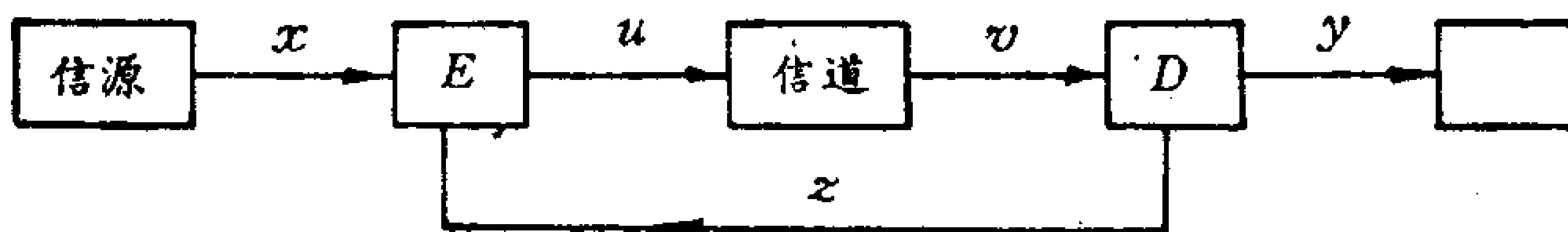


图5.12 有反馈信号的通信系统

对这些模型的编码问题的讨论内容十分丰富。在本书中我们重点介绍两个多用户信息论的典型模型，即多重信源中的SWC模型与二址信道模型。

§5.3 多重信源的编码定理

1.SWC模型的编码问题

从§5.2图5.1的SWC-模型可以看出，该通信系统有二个相关信源构成。我们记之为

$$\boldsymbol{S}^n=[\boldsymbol{X}_1^n\times\boldsymbol{X}_2^n,\ p^n(x_1^n,\ x_2^n)],\ n=1,2,3,\cdots, \tag{3.1}$$

其中$p^n(x_1^n,x_2^n)$为$\boldsymbol{X}_1^n\times\boldsymbol{X}_2^n$上的联合概率分布。我们称(3.1)的$\boldsymbol{S}^n$为一个二重相关信源序列。由§5.2图5.1的要求，该系统的编码函数（f^n，g^n)应这样规定。其中$f^n(x_1^n,x_2^n)=(f_1^n(x_1^n),f_2^n(x_2^n))$，而$f_k^n$分别为映射

$$f_k^n: \boldsymbol{X}_k^n \to \boldsymbol{M}_k^n = \{1,2,\cdots, M_k^n\}, \quad k=1, 2. \quad (3.2)$$

如我们记M_1^n与M_2^n中的元分别为i, j，则译码函数$g^n(i, j)=(g_1^n(i, j), g_2^n(i, j))$为映射

$$g_k^n: \boldsymbol{M}_1^n \times \boldsymbol{M}_2^n \to \boldsymbol{Y}_k^n, \quad k=1, 2, \quad (3.3)$$

其中$\boldsymbol{Y}_k^n = \boldsymbol{X}_k^n$, $k=1, 2$。

由（3.1）—（3.3）给定的通信系统是SWC-模型的严格数学描述。当$\boldsymbol{S}^n$，(f^n, g^n）给定后，$\boldsymbol{X}_1^n \times \boldsymbol{X}_2^n \times \boldsymbol{M}_1^n \times \boldsymbol{M}_2^n \times \boldsymbol{Y}_1^n \times \boldsymbol{Y}_2^n$上的联合概率分布确定。相应的随机变量为（$X_1^n$, X_2^n, U_1^n, U_2^n, Y_1^n, Y_2^n）。它们的联合概率分布为

$$p^n(x_1^n, x_2^n, u_1^n, u_2^n, y_1^n, y_2^n)$$
$$= p^n(x_1^n, x_2^n) f^n[(u_1^n, u_2^n)/(x_1^n, x_2^n)] g^n[y_1^n, y_2^n)/(u_1^n, u_2^n)], \quad (3.4)$$

其中$p^n(x_1^n, x_2^n)$由（3.1）给定。而

$$f^n[(u_1^n, u_2^n)/(x_1^n, x_2^n)] = \begin{cases} 1, \text{如果} u_k^n = f_k^n(x_k^n), k=1,2, \\ 0, \text{否则。} \end{cases} \quad (3.5)$$

$$g^n[(y_1^n, y_2^n)/(u_1^n, u_2^n)] = \begin{cases} 1, \text{如果} y_k^n = g_k^n(u_1^n, u_2^n), k=1, 2, \\ 0, \text{否则.} \end{cases} \quad (3.6)$$

定义1 对由（3.1）给定的二重信源序列$\boldsymbol{S}^n$，我们称$R=(R_1, R_2)$为SWC-系统的一个可达速率对，如对任何$\varepsilon>0$，只要

$$M_k^n \geqslant \exp_2[n \cdot R_k \cdot (1+\varepsilon)], \quad (3.7)$$

那么当n充分大，就有一组由（3.2），（3.3）确定的编码（f^n, g^n），使

$$P_r\{(X_1^n, X_2^n) \neq (Y_1^n, Y_2^n)\} < \varepsilon \quad (3.8)$$

成立。我们记$\boldsymbol{R}$为$\boldsymbol{S}^n$的全体可达速率对，且称之为可达速率区

域。

多重信源的编码问题是找出多重信源的可达速率区域。

2. SWC-模型的编码定理

为求出SWC-模型的可达速率区域，我们讨论无记忆的信源序列。这时（3.1）的$\boldsymbol{S}^n$满足条件：

$$\boldsymbol{X}_k^n=\boldsymbol{X}_{k,1}\times\boldsymbol{X}_{k,2}\times\cdots\times\boldsymbol{X}_{k,n},\quad k=1,2,\qquad(3.9)$$

其中$\boldsymbol{X}_{k,i}=\boldsymbol{X}_k$，$i=1,2,\cdots,n$，$k=1,2$。而

$$p^n(x_1^n,x_2^n)=\prod_{i=1}^{n}p(x_{1,i},x_{2,i}).\qquad(3.10)$$

这时$\boldsymbol{S}=[\boldsymbol{X}_1\times\boldsymbol{X}_2,\ p(x_1,x_2)]$为$\boldsymbol{S}^n$的二重生成信源。以下记由$p(x_1,x_2)$确定的随机变量为$(X_1,X_2)$。

定理1 如果$\boldsymbol{S}^n$为由（3.9），（3.10）给定的二重无记忆信源序列，由$\boldsymbol{S}$生成。这时$\boldsymbol{S}^n$的SWC-可达速率区域为

$$\boldsymbol{R}=\{R=(R_1,R_2):\ R_1\geqslant H(X_1),$$
$$R_2\geqslant H(X_2),\quad R_1+R_2\geqslant H(X_1,X_2)\},\qquad(3.11)$$

其中$H(X_1)$，$H(X_2)$，$H(X_1,X_2)$分别为X_1,X_2的熵与联合熵。

证明：对这个定理的正命题我们用随机码的方法来证。对它的逆命题同样可用法诺不等式证。我们证明它的正命题。对任何(3.11)中的$R=(R_1,R_2)$与任何$\varepsilon>0$，我们记

$$M_k^n=\exp_2[n\cdot R_k\cdot(1+\varepsilon)],\ k=1,2.\qquad(3.12)$$

我们构造编码如下。

(1) 随机编码(f_1^{n*},f_2^{n*})为这样定义：

对每个$x_k^n\in\boldsymbol{X}_k^n$，取$f_k^{n*}(x_k^n)$为独立同分布的随机变量，且在

$$\boldsymbol{M}_k^n=\{1,2,3,\cdots,M_k^n\},\ k=1,2,\qquad(3.12')$$

上取均匀分布，也就是对任何$x_k^n \in \boldsymbol{X}_k^n$，$j \in \boldsymbol{M}_k^n$，总有

$$P_r\{f_k^{n*}(x_k^n)=j\}=1/M_k^n \tag{3.13}$$

成立，且

$$\{f_1^{n*}(x_1^n),\ x_1^n \in \boldsymbol{X}_1^n\},\ \{f_2^{n*}(x_2^n),\ x_2^n \in \boldsymbol{X}_2^n\}$$

是两组相互独立的随机变量。这时，f_k^{n*} 的一个样本值f_k^n决定$\boldsymbol{X}_k^n$上的一个分割，我们记之为

$$\{\boldsymbol{A}_{k,j}^n,\ j=1,\ 2,\ \cdots,\ M_k^n\}, \tag{3.14}$$

它们互不相交且并为$\boldsymbol{X}_k^n$，且有

$$f_k^n(x_k^n)=j,\ \text{当}x_k^n \in \boldsymbol{A}_{k,j}^n\text{时},\ k=1,\ 2. \tag{3.15}$$

(2) 译码函数g^n的定义为：

如果信源序列$\boldsymbol{S}^n$，编码函数$f^n=(f_1^n,\ f_2^n)$给定，其中f^n 为f^{n*}的一个样本值。由编码函数$f^n=(f_1^n,\ f_2^n)$ 确定$\boldsymbol{X}_1^n$，$\boldsymbol{X}_2^n$的两个分割 (3.14)，使 (3.15) 成立。这时对任何$(i,\ j) \in \boldsymbol{M}_1^n \times \boldsymbol{M}_2^n$，我们构造译码函数

$$g^n(i,\ j)=(x_1^n,\ x_2^n), \tag{3.16}$$

如果 $(x_1^n,\ x_2^n)$满足以下条件：

(i) $(x_1^n,\ x_2^n) \in \boldsymbol{A}_{1,i}^n \times \boldsymbol{A}_{2,j}^n$；

(ii) 有关系式

$$h_k(x_k^n)<n \cdot R_k \cdot (1+\varepsilon/2),\ k=0,\ 1,\ 2, \tag{3.17}$$

成立，(3.17)中有关记号是这样的，$R_0=R_1+R_2$，$x_0^n=(x_1^n,\ x_2^n)$，而

$$h_k(x_k^n)=-\log p_k^n(x_k^n),\ k=0,\ 1,\ 2, \tag{3.18}$$

其中 $p_0^n(x_0^n)=p^n(x_1^n,\ x_2^n)$，而

$$p_1^n(x_1^n)=\sum_{x_2^n} p^n(x_1^n,\ x_2^n),$$

$$p_2^n(x_2^n)=\sum_{x_1^n} p^n(x_1^n,\ x_2^n).$$

(iii) (x_1^*, x_2^*)为唯一使条件(1)，(2) 成立的向量对。

如果条件(i)，(ii)与(iii)不能同时成立，则取$g^n(i, j)$ 为$\mathbf{X}_1^n \times \mathbf{X}_2^n$中的任意一个向量。

(3) 随机编码与译码的平均误差概率。

以下记$\mathbf{G}_0^n = \mathbf{G}_0^n(\varepsilon)$为使条件(3.17) 成立的全体$(x_1^n, x_2^n)$，而$\mathbf{G}_1^n = \mathbf{G}_1^n(\varepsilon)$为$\mathbf{G}_0^n(\varepsilon)$的余集，也就是全体不在 $\mathbf{G}_0^n(\varepsilon)$ 中的(x_1^n, x_2^n)。对固定的编码 (f^n, g^n)，我们定义平均误差概率为

$$p_e(f^{n*}, g^{n*}) = P_r\{(X_1^n, X_2^n) \neq g^{n*}[f^{n*}(X_1^n, X_2^n)]\},$$

其中 (f^{n*}, g^{n*}) 为由(1)，(2) 给定的编码，当f^{n*}给定为f^n时g^{n*}就确定为 (2)的g^n。由 (f^{n*}, g^{n*}) 的定义可知，

$$p_e(f^{n*}, g^{n*}) \leqslant P_r\{(X_1^n, X_2^n) \in \mathbf{G}_1^n\} + P_r\{(X_1^n, X_2^n) \in \mathbf{G}_0^n \text{ 且 } (X_1^n, X_2^n) \neq g^{n*}[f^{n*}(X_1^n, X_2^n)]\} \tag{3.19}$$

成立。因为

$$p^n(x_1^n, x_2^n) = \prod_{t=1}^{n} p(x_{1,t}, x_{2,t})$$

是无记忆信源的乘积分布，其中$x_k^n = (x_{k,1}, \cdots, x_{k,n})$，$k=1$，2。因此由大数定律可得极限式

$$\frac{1}{n}h(x_1^n, x_2^n) = -\frac{1}{n}\sum_{t=1}^{n}\log p(x_{1,t}, x_{2,t}) \to H(X_1, X_2),$$

$$\frac{1}{n}h_k(x_k^n) = -\frac{1}{n}\sum_{t=1}^{n}\log p_k(x_{k,t}) \to H(X_k), \quad k=1, 2,$$

其中“$\to$”为以概率收敛，而$p_1(x_1) = \sum_{x_2} p(x_1, x_2)$，$p_2(x_2) = \sum_{x_1} p(x_1, x_2)$。这时对任何$\varepsilon > 0$，只要$n$充分大，就有

$$P_r\{(X_1^n,\ X_2^n)\in \boldsymbol{G}_1^n(\varepsilon)\}\leqslant \varepsilon/2 \tag{3.20}$$

成立。我们现在估计（3.19）式中不等式右边第二项的值，对该项的值我们记之为$p_{e,0}(f^{n*},\ g^{n*})$。

首先由编码函数f^n的定义可知，对任何$(x_1^n, x_2^n)\in \boldsymbol{G}_0^n(\varepsilon)$，总有一对整数$(i,\ j)\in \boldsymbol{M}_1^n\times \boldsymbol{M}_2^n$使$f_1^n(x_1^n)=i, f_2^n(x_2^n)=j$成立，因此

$$\begin{aligned}
p_{e,0}(f^{n*},\ g^{n*})\leqslant &\sum_{(x_1^n, x_2^n)\in G_0^n} p^n(x_1^n,\ x_2^n)\cdot P_r\{\text{有一对}(x_1^{n\prime},\ x_2^{n\prime})\in \boldsymbol{G}_0^n,\ \text{使}\\
&(x_1^{n\prime},\ x_2^{n\prime})\neq(x_1^n,\ x_2^n),\ \text{而} f_k^{n*}(x_k^{n\prime})=f_k^{n*}(x_k^n),\\
&k=1,\ 2\}\\
\leqslant &\sum_{x_1^n,\ x_2^n} p^n(x_1^n,\ x_2^n)\cdot\left[\sum_{z=1}^{3} p_{e,z}(x_1^n,\ x_2^n)\right].
\end{aligned} \tag{3.21}$$

在（3.21）中$p_{e,z}(x_1^n,\ x_2^n)$的定义为

$$\begin{aligned}
&p_{e,1}(x_1^n,\ x_2^n)=p_{e,1}(x_1^n)\\
&=P_r\{\text{有一个}x_1^{n\prime}\in \boldsymbol{G}_{0,1}^n,\ \text{使}\ x_1^{n\prime}\neq x_1^n,\ \text{而}\\
&f_1^{n*}(x_1^{n\prime})=f_1^{n*}(x_1^n)\},\\
&p_{e,2}(x_1^n,\ x_2^n)=p_{e,2}(x_2^n)\\
&=P_r\{\text{有一个}x_2^{n\prime}\in \boldsymbol{G}_{0,2}^n,\ \text{使}x_2^{n\prime}\neq x_2^n,\ \text{而}\\
&f_2^{n*}(x_2^{n\prime})=f_2^{n*}(x_2^n)\},\\
&p_{e,3}(x_1^n,\ x_2^n)=P_r\{\text{有一对}\ (x_1^{n\prime},\ x_2^{n\prime})\in \boldsymbol{G}_0^n,\ \text{使}\\
&x_1^{n\prime}\neq x_1^n,\ x_2^{n\prime}\neq x_2^n,\\
&\text{而且}\ f_1^{n*}(x_1^{n\prime})=f_1^{n*}(x_1^n),\ f_2^{n*}(x_2^{n\prime})=f_2^{n*}(x_2^n)\},
\end{aligned}$$

其中

$$\boldsymbol{G}_{0,k}^n \equiv \{x_k^n:\ h_k(x_k^n)<nR_k(1+\varepsilon/2)\},\ k=0,1,2, \tag{3.22}$$

$h_k(x_k^n)$，$k=0,1,2$，为(3.18)定义，而$R_0=R_1+R_2$。这时

$$\boldsymbol{G}_0^n=\boldsymbol{G}_{0,0}^n\cap\boldsymbol{G}_{0,1}^n\cap\boldsymbol{G}_{0,2}^n.$$

我们现在估计$p_{e,s}(x_1^n,\ x_2^n)$，$s=1,2,3$，的值．由概率的性质，$\boldsymbol{G}_{0,1}^n$与f_1^{n*}的定义，可得

$$\begin{aligned}&p_{e,1}(x_1^n,\ x_2^n)\leqslant(\|\boldsymbol{G}_{0,1}^n\|-1)\cdot(1/M_1^n)\\&\leqslant\exp_2\{n\cdot[H(X_1)\cdot(1+\varepsilon/2)-R_1\cdot(1+\varepsilon)]\}\\&\leqslant\exp_2(-n\cdot R_1\cdot\varepsilon/2)\to0.\end{aligned}\qquad(3.23)$$

而(3.23)的不等式是由以下关系推出：

(i) 由可达速率区域$\boldsymbol{R}$的定义 (3.11)式，有$R_1\geqslant H(X_1)$成立。

(ii) 由$\boldsymbol{G}_{0,1}^n$的定义可知，对任何$x_1^n\in\boldsymbol{G}_{0,1}^n$，有

$$h(x_1^n)\leqslant n\cdot R_1\cdot(1+\varepsilon/2)\text{或}p_1^n(x_1^n)\geqslant\exp_2[-n\cdot R_1\cdot(1+\varepsilon/2)]$$

成立，因此有

$$\|\boldsymbol{G}_{0,1}^n\|\leqslant\exp_2[n\cdot R_1\cdot(1+\varepsilon/2)].$$

(iii) 由f_1^{n*}的定义可得，对任何$x_1^n\neq x_1^{n\prime}$，必有

$$P_r\{f_1^{n*}(x_1^n)=f_1^{n*}(x_1^{n\prime})\}=1/M_1^n=\exp_2[-n\cdot R_1\cdot(1+\varepsilon)]$$

成立．由(i)—(iii)即可推出(3.22)式。

同理可得$p_{e,k}(x_1^n,\ x_2^n)\to0$成立，$k=2,3$。由(3.21)即得，当$n$充分大时，有$p_{e,0}(f_1^{n*},g_2^{n*})\leqslant\varepsilon/2$成立。由(3.19)与(3.20)即得，当$n$充分大时，有$p_e(f_1^{n*},g_2^{n*})\leqslant\varepsilon$成立，因此$R=(R_1,R_2)$为可达速率。定理的正命题得证。

对该定理逆命题，即对任何$\boldsymbol{R}$区域外的向量$R=(R_1,R_2)$一定不是可达速率，对此命题的证明思路与§3.3定理1的逆命题相似，对此不再详述。

由SWC-模型的编码要求及定理1的结论与证明，可以看到

多重信源编码理论的基本特征.对§5.2中的各种多重信源同样可给出它的编码要求与可达速率区域，对此不作一一详述。

§5.4 多址信道的容量区域与编码定理

1.二址信道的编码问题

二址信道的通信模型在§5.2图5.7中已经描述。我们记

$$\boldsymbol{C}=[\boldsymbol{U}_1\times\boldsymbol{U}_2,\ p(w/u_1,u_2),\boldsymbol{W}] \qquad (4.1)$$

为一个二址信道，其中$\boldsymbol{U}_1$，$\boldsymbol{U}_2$为信道的二个输入信号字母表，$\boldsymbol{W}$ 为信道的输出信号字母表，相应的输入、输出信号为 u_1，u_2; w，而$p(w/u,v)$为条件转移概率。我们同样称

$$\boldsymbol{C}^n=[\boldsymbol{U}_1^n\times\boldsymbol{U}_2^n,\ p^n(w^n/u_1^n,u_2^n),\ \boldsymbol{W}^n],n=1,2,\cdots, \qquad (4.2)$$

为二址信道的信道序列，其中$\boldsymbol{U}_1^n$，$\boldsymbol{U}_2^n$为二址信道的输入信号字母表序列，$\boldsymbol{W}^n$为信道的输出信号字母表序列，相应的输入、输出信号为u_1^n,u_1^n; w^n，而$p^n(w^n/u^n,v^n)$为条件转移概率. 以下记

$$\boldsymbol{S}^n=(\boldsymbol{S}_1^n,\boldsymbol{S}_2^n)=[\boldsymbol{M}_1^n\times\boldsymbol{M}_2^n,p^n(x_1^n,x_2^n)] \qquad (4.3)$$

为一个二址信道的二个输入消息信源，其中

$$\boldsymbol{M}_k^n=\{1,2,\cdots,M_k^n\},\ k=1,2 \qquad (4.4)$$

$$p^n(x_1^n,x_2^n)=p_1^n(x_1^n)p_2^n(x_2^n),\ p_k^n(x_k^n)=1/M_k^n,k=1,2. \qquad (4.4')$$

这时编码序列为$(f^n,g^n)=[(f_1^n,f_2^n),(g_1^n,g_2^n)]$，其中$f_k^n,g_k^n$，$k=1,2$,分别为映射

$$f_k^n:\ \boldsymbol{M}_k^n\to\boldsymbol{U}_k^n,\ g_k^n:\ \boldsymbol{W}^n\to\boldsymbol{M}_k^n,\ k=1,2. \qquad (4.5)$$

如果(4.2)—(4.5) 的二址通信系统的信源$\boldsymbol{S}^n$，信道 $\boldsymbol{C}^n$ 与编码

(f^n,g^n)给定，那么相应的信源、信道的输入、输出随机变量确定，我们记之为X_1^n,X_2^n,U_1^n,U_2^n与W^n，Y_1^n，Y_2^n，其中Y_k^n为在$\mathbf{M}_k^n$，$k=1,2$上取值的随机变量，相应的联合概率分布为

$$\begin{aligned}
&p(x_1^n,x_2^n,u_1^n,u_2^n,w^n,y_1^n,y_2^n)\\
&=P_r\{(X_1^n,X_2^n,U_1^n,U_2^n,W^n,Y_1^n,Y_2^n)\\
&=(x_1^n,x_2^n,u_1^n,u_2^n,w^n,y_1^n,y_2^n)\}\\
&=p^n(x_1^n,x_2^n)f^n(u_1^n,u_2^n/x_1^n,x_2^n)\\
&\quad\cdot p^n(w^n/u_1^n,u_2^n)g^n(y_1^n,y_2^n/w^n),
\end{aligned}\tag{4.6}$$

其中

$$f^n(u_1^n,u_2^n/x_1^n,x_2^n)=\begin{cases}1, & 如果(u_1^n,u_2^n)=f^n(x_1^n,x_2^n),\\ 0, & 否则。\end{cases}\tag{4.6'}$$

$$g^n(y_1^n,y_2^n/w^n)=\begin{cases}1, & 如果(y_1^n,y_2^n)=g^n(w^n),\\ 0, & 否则。\end{cases}\tag{4.6''}$$

相应的误差概率为

$$p_e(f^n,g^n)=P_r\{(X_1^n,X_2^n)\neq(Y_1^n,Y_2^n)\}.\tag{4.7}$$

定义1 称$R=(R_1,R_2)$为二址信道$\mathbf{C}^n$的一个可达速率，如果对任何$\varepsilon>0$，当我们取(4.3)的$M_k^n=\exp_2[nR_k(1-\varepsilon)]$时，只要$n$充分大，就有一列(4.5)的编码序列$(f^n,g^n)$，使$p_e(f^n,g^n)<\varepsilon$成立。

同样称可达速率向量的全体为可达速率区域。二址信道的编码问题就是寻找它的可达速率区域。

2.容量区域与编码定理

与§3.2，§3.3的情形相同，为求通信系统的可达速率区域，我们仍借助于交互信息。与§3.2不同之处是多用户信道的容量是一个多维向量的区域。我们先考虑二址信道的情形，对

(4.1)的信道$\boldsymbol{C}=[\boldsymbol{U}_1^n\times\boldsymbol{U}_2^n,\ p(w^n/u_1,u_2),W]$，我们称

$$p(u_1,u_2)=p_1(u_1)\cdot p_2(u_2),\quad (u_1,u_2)\in\boldsymbol{U}_1\times\boldsymbol{U}_2 \tag{4.8}$$

为二址信道的入口分布，其中$p_k(u_k)$为$\boldsymbol{U}_k$上的概率分布。我们记全体(4.8) 型的二址信道的入口分布为$\boldsymbol{P}(\boldsymbol{U}_1\times\boldsymbol{U}_2)$，相应的元为$p(\boldsymbol{U}_1\times\boldsymbol{U}_2)$。

如果二址信道$\boldsymbol{C}$与它的入口分布$p(\boldsymbol{U}_1\times\boldsymbol{U}_2)$给定，那么信道的输入、输出随机变量$U_1$，$U_2$，$W$ 确定，它们的联合概率分布为

$$\begin{aligned}p(u_1,u_2,w)&=P_r\{(U_1,U_2,W)=(u_1,u_2,w)\}\\&=p(u_1,u_2)p(w/u_1,u_2)\\&=p_1(u_1)p_2(u_2)p(w/u_1,u_2)。\end{aligned} \tag{4.9}$$

而相应的交互信息为

$$I((U_1,U_2);W)=\sum_{u_1,u_2,w}p(u_1,u_2,w)\log\frac{p(u_1,u_2,w)}{p(u_1,u_2)q(w)}, \tag{4.10}$$

$$I(u_2;W/U_1)=\sum_{u_1,u_2,w}p(u_1,u_2,w)\log\frac{p(u_2,w/u_1)}{p_2(u_2)p(w/u_1)}, \tag{4.10$'$}$$

$$I(U_1;W/U_2)=\sum_{u_1,u_2,w}p(u_1,u_2,w)\log\frac{p(u_1,w/u_2)}{p_1(u_1)p(w/u_2)}, \tag{4.10$''$}$$

其中 $p_1(u_1)$，$p_2(u_2)$，$p(u_1,w/u_2)$，$p(u_2,w/u_1)$，$q(w/u_1)$，$q(w/u_2)$均为$p(u_1,u_2,w)$的边际分布与条件分布．以下记

$$\begin{aligned}&I_0=I[(U_1,U_2);W],\ I_1=I(U_1;W/U_2),\\&I_2=I(U_2;W/U_1)。\end{aligned} \tag{4.11}$$

我们定义

$$\mathbf{R}[p(\mathbf{U}_1\times\mathbf{U}_2)]=\{R=(R_1,R_2):\ 0\leqslant R_k\leqslant I_k,\ k=0,1,2\},\qquad(4.12)$$

其中I_k，$k=0,1,2$，为(4.11)定义，而$R_0=R_1+R_2$。

定义1 对(4.1)的二址信道$\mathbf{C}$的容量区域$\mathbf{R}(\mathbf{C})$定义为集合$\mathbf{R}[p(\mathbf{U}_1\times\mathbf{U}_2)]$对全体不同的$p(\mathbf{U}_1\times\mathbf{U}_2)\in\mathbf{P}(\mathbf{U}_1\times\mathbf{U}_2)$的并。

对(4.2)的二址信道序列$\mathbf{C}^n$，$n=1,2,\cdots$，我们同样可定义它的容量区域$\mathbf{R}(\mathbf{C}^n)$，它为集合$\mathbf{R}[p(\mathbf{U}_1^n\times\mathbf{U}_2^n)]$对全体不同的$p(\mathbf{U}_1^n\times\mathbf{U}_2^n)\in\mathbf{P}(\mathbf{U}_1^n\times\mathbf{U}_2^n)$的并，其中$\mathbf{P}(\mathbf{U}_1^n\times\mathbf{U}_2^n)$为$\mathbf{U}_1^n\times\mathbf{U}_2^n$上的全体乘积型（也就是具有(4.8)型）的概率分布，而

$$\mathbf{R}[p(\mathbf{U}_1^n\times\mathbf{U}_2^n)]=\{R=(R_1,R_2):\ 0\leqslant R_k\leqslant I_k^n/k=0,1,2\},\qquad(4.12')$$

其中

$$I_0^n=I[(U_1^n,U_2^n);W^n],\quad I_1^n=I(U_1^n;W^n/U_2^n),$$
$$I_2^n=I(U_2^n;W^n/U_1^n)\qquad(4.11')$$

为相应随机变量的交互信息或条件交互信息。

对二址信道同样可定义它们的无记忆性。如果(4.2)的$\mathbf{C}^n$满足条件

(1) 它的字母表分别为n-维乘积型向量空间，这时

$$\mathbf{U}_k^n=\prod_{t=1}^{n}\mathbf{U}_{k,t},\ k=1,2;\quad \mathbf{W}^n=\prod_{t=1}^{n}\mathbf{W}_t,\qquad(4.13)$$

其中$\mathbf{U}_{k,t}=\mathbf{U}_k$，$k=1,2$；$\mathbf{W}_t=\mathbf{W}$。

(2) 转移概率分布

$$p^n(w^n/u_1^n,u_2^n)=\prod_{t=1}^{n}p(w_t/u_{1,t},u_{2,t})\qquad(4.13')$$

对任何$u_1^n=(u_{1,1},u_{1,2},\cdots,u_{1,n})$，$u_2^n=(u_{2,1},u_{2,2},\cdots,u_{2,n})$，$w^n=(w_1,w_2,\cdots,w_n)$成立。这时我们称$\mathbf{C}^n$为由$\mathbf{C}$生成的无记忆二址信道序列。

定理1 如果$\mathbf{C}^n$为由$\mathbf{C}$生成的无记忆二址信道序列，那么它们的信道容量区域为$\mathbf{R}(\mathbf{C}^n)=n\cdot\mathbf{R}(\mathbf{C})$，其中

$$n\cdot\mathbf{R}=\{(n\cdot R_1,n\cdot R_2):\ R=(R_1,R_2)\in\mathbf{R}\}.$$

该定理的证明与§3.2的无记忆信道容量的性质定理的证明相似，在此只要进一步利用条件交互信息的性质就可证明定理。

定理2 如果$\mathbf{C}^n$为由$\mathbf{C}$生成的无记忆二址信道序列，那么$\mathbf{C}^n$的可达速率区域就是$\mathbf{C}$的信道容量区域。

对该定理的证明同样可用“随机编码”与“信息门限”译码的方法来证。对此我们也不再详述了。

3.多用户信息论的其它问题

多址信道是二址信道的形式推广，一个K-址信道序列可记为

$$\mathbf{C}_k^n=[\mathbf{U}_1^n\times\mathbf{U}_2^n\times\cdots\times\mathbf{U}_k^n,p(w/u_1^n,u_2^n,\cdots,u_k^n),\mathbf{W}^n],\quad n=1,2,3,\cdots \tag{4.14}$$

对此我们可同样定义它的编码函数、可达速率区域与容量区域。且对它们的编码定理可作类似的推广。

对其它的多用户通信系统（如广播信道、有边信息信道、有反馈信号信道等），它们的可达速率区域与容量区域的关系在许多情形不具有等价关系。因此，它们的编码理论要比本章中的SWC-模型与二址信道模型复杂得多，在这些模型中，人们只能找到若干区域的上、下界关系，从而构成多用户信息论中的一系列未解决问题。

第六章　信息量在信息科学的其它分支中的应用

由于信息科学的最新发展，信息的度量在更多的学科领域中得到应用与发展．典型的问题有如“信息量在密码学中的应用”；“信息量与计算机复杂性理论及与分形几何的关系”；“信息统计理论”等．由于这些分支涉及许多专门知识，在本章中我们只作简单的描述，通过这些描述可以看到信息度量在更多领域中的作用．

§6.1　信息量在密码学中的应用

1.密码系统的模型与要素

一个密码系统与一个通信系统相似．它同样由信源、信道、加密、解密运算构成，同时它又有“密钥源”，“干扰源”与“分析者”存在．它的框图为图6.1所示．

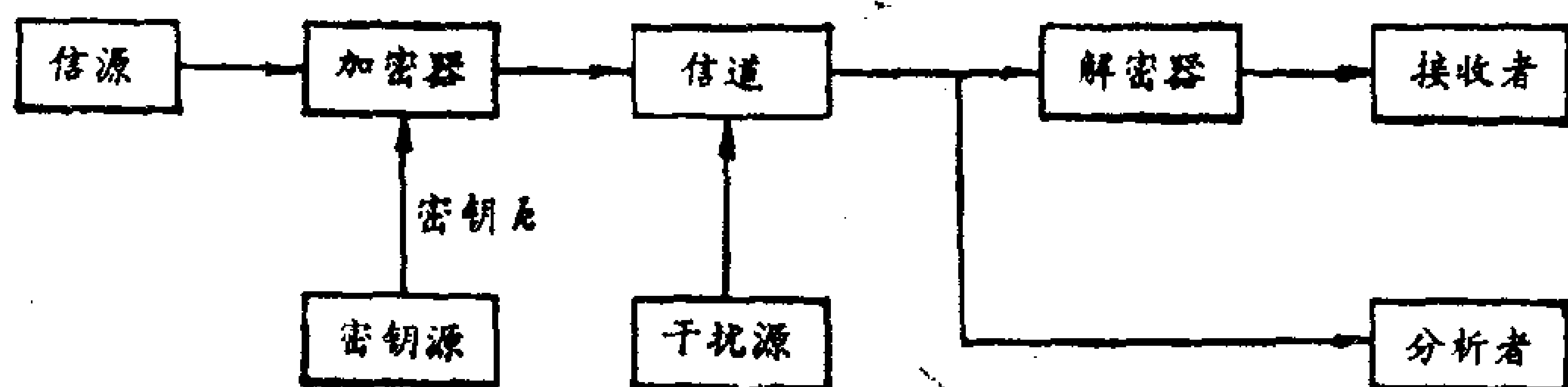


图6.1 密码系统结构框图

其中信源、信道的描述与§2.1相同，我们仍记为

$$\boldsymbol{S}=[\boldsymbol{X},p(x)],\quad \boldsymbol{C}=[\boldsymbol{U},p(v/u),\boldsymbol{V}],\qquad (1.1)$$

为了简单起见，在密码系统中常取信道$\boldsymbol{C}$为无噪声的，这时$\boldsymbol{V}=\boldsymbol{U}$，且$p(u/u)=1$，$p(v/u)=0$，当$v\neq u$时。

加密运算一般是一个带参数的映射

$$T_k(x):\ \boldsymbol{X}\to\boldsymbol{U},\ k\in\boldsymbol{K},\qquad (1.2)$$

其中$\boldsymbol{K}$为加密运算的参数空间，我们称之为密钥空间。对密钥的选取一般也是随机的，以下记$q(k)$，$k\in\boldsymbol{K}$为密钥选取的概率分布，我们称

$$\boldsymbol{E}=\{[\boldsymbol{X}\times\boldsymbol{K},\ p(x,k)],\ T_k\in\boldsymbol{K}\},\qquad (1.3)$$

为一个密码系统，其中$p(x,k)$为明文与密钥的联合概率分布。在密码系统中，明文与密钥的选取一般是相互独立的，这时$p(x,k)=p(x)q(k)$。

在密码系统中，信源又称为明文信源(或简称为明文)，明文经过加密运算后变为密文。对密文一般不再作其他的加密措施。在框图6.1中，解密运算是加密运算的逆运算，我们记之为$T_k^{-1}(u)$，这时$T_k^{-1}(u)$为一个从$\boldsymbol{U}$到$\boldsymbol{Y}=\boldsymbol{X}$的映射。我们称$y=T_k^{-1}(u)$为明文的复制。加密与解密运算是密码系统的核心部分，近代密码学就是以密钥的选取为保密学的基础。因此，人们常把加密、解密运算称为密码体制。

如我们记X，K，U，Y分别为明文、密钥、密文与明文复制的随机变量，那么它们的联合概率分布为

$$p(x,k,u,y)=P_r\{(X,K,U,Y)=(x,k,u,y)\}$$
$$=p(x)q(k)T(u/x,k)T^{-1}(y/u,k), \quad (1.4)$$

其中 $$T(y/x,k)=\begin{cases}1, & \text{如果}y=T_k(x),\\ 0, & \text{否则}.\end{cases}$$

$$T^{-1}(y/x,k)=\begin{cases}1, & \text{如果}y=T_k^{-1}(u),\\ 0, & \text{否则}.\end{cases}$$

密码系统的破坏者称为“对手”，对手的类型有“窃听分析型”与“干扰型”。其中窃听分析型是通过对密文的收集与分析来了解明文，而干扰型是对明文或密文数据进行干扰，使用户自己不能使用通信系统或产生错误的决策。防窃听分析型的密码系统为保密系统，而防干扰型的密码系统为数据安全系统。这两种系统在数学构造上无原则的差别，但它们的具体特征在计算机运作中是不相同的。

近代密码学是计算机与信息科学中重要分支，除了在传统的政府、军事领域之外，在企事业、金融、新闻乃至家庭个人生活中都有许多应用。它的内容、方法也很丰富。典型的密码体制有如DES体制，公钥体制等，其中*DES*体制是美国的一种国家标准加密体制，而公钥体制是一种把加密密钥与解密密钥区分且可把加密密钥公开的密码体制，它是一种应用广泛的近代密码体制。在本节中，我们只介绍与信息量有关的内容。

2.密码系统的评价标准

利用信息量可对密码系统进行一系列的好坏评价。主要指标有以下几项。

(1) 密码系统的完善性

上文已经说明，一个窃听者是通过密文来分析明文的，也就是，窃听者是通过密文来获取明文的信息。因此，一个理想的密码系统是它的密文不含任何明文的信息。

定义1 一个密码系统E被称为是完善的(perfect) 密码系统，如果明文X与密文Y是相互独立的随机变量,也就是它们的交互信息$I(X;Y)=0$。

一个完善的密码系统是理想的密码系统，任何分析者都不能由密文得到任何关于明文的信息。但另一方面，在实际的密码系统中，由于密钥空间的有限性及通信过程的无限性（也就是通信数据可无限延长),因此，对任何一个有限的密钥空间，在一个长期使用的密码系统中要永远保持完善性是不可能的。因此我们必需考虑一个实际上不可破的密码体制。一个实际上不可破的密码体制必需具备以下两个条件，即

(i) 密文关于明文的条件熵 $H(X/U)$ 必需足够大。这就是分析者在获知密文U后，明文X的不肯定性必需很大。

(ii) 分析者在未知密钥k的情形下,由密文u求明文x的运算是非易计算的。非易计算的术语是计算复杂性理论中的术语，它的严格定义就不叙述了。

(2) 密码体制的剩余度分析

在实际通信中，分析者有可能同时获得部分明文与密文，例如若干历史资料的公开可使分析者同时获得这些资料的明文与密文，一个可靠的密码体制必需保证在这些资料泄露后的安全性。这就要求密码系统的条件熵$H(X/U,X_0,U_0)$ 足够大，其中X_0,U_0为历史的明文与密文资料，而称$H(X/U,X_0,U_0)$为剩余度

利用信息量可对各种密码体制的完善性进行分析与对$H(X$

$/U)$，$H(X/U,X_0,U_0)$进行计算．对上述密码系统我们只给出了一个静态模形，对动态情形必需引进序列模型．因此本节的描述只是定性的说明．

（3）其它指标有如：理论或实际的不可破性；设备与操作的简便性；与计算机的匹配性等等．因为这些条件与信息量无关，所以在本书中不作详述了．

§6.2 信息量与计算机复杂性理论与分形几何的关系

1.算法信息论与柯尔莫各洛夫复杂性

80年代初，人们对信息量的改进提出了一条新的途径，它与计算机复杂性理论密切相关，因此人们把它称为“算法信息论”.算法信息论的基本观点如下：

（1）算法信息论的基本观点之一是认为，仙农信息量的一个主要不合理性是不考虑每个事件本身所具有的信息．因为在仙农熵的定义中只考虑每个事件的概率分布．事实上，在信息处理中，每个事件不仅有出现的概率有所不同，而且每个事件本身可能具有不同的信息．例如，在一篇文章中，有的词句特别重要，而有的词句则无关紧要；又如在数值计算中，对99^9与9^{9^9}两个数，它们出现的符号完全相同（都是三个“9”字），但它们的计算量有很大的差别．

因此，算法信息论的第一个基本观点是必需考虑每个事件本身所具有的信息．

（2）为考虑不同事件本身所具有的信息，必需有一种统一

的信息度量标准。因为客观事物千差万别，要建立这种统一的信息度量是十分困难的。算法信息论的第二个基本观点是对这种统一的信息度量标准应以计算复杂性为基础来进行考虑。算法信息论认为，计算复杂性就是信息的度量。

这样算法信息论就把信息度量与计算复杂性溶为一体。因此，算法信息论是信息论与计算机科学的结合点。

算法信息论中关于信息度量的核心是柯尔莫各洛夫复杂度。关于柯尔莫各洛夫复杂度的严格定义与性质的讨论要涉及到递归函数理论，因此我们只能概述它的基本思路。

如果我们记x为计算机要计算的结果数据，为实现对x的计算（使x成为计算机的输出），我们必需设计一个程序P，使这个程序P在特定或通用的计算机T中进行运算，使运算结果$T(P)=x$。为实现“$T(P)=x$”这个目标，我们可以设计许多程序，其中最短的程序长度就为柯尔莫各洛夫复杂度。因此，柯尔莫各洛夫复杂度的定义为

$$K(x)=\min\{l(P): T(P)=x\}, \qquad (2.1)$$

其中$l(P)$为程序长度。

(2.1) 的严格定义要用递归函数来叙述。因为上述柯尔莫各洛夫复杂性是用程序长度来计算，因此在计算复杂性理论中又把它称为“描述复杂性”。

2.分形几何与豪斯道夫维数

传统的几何维数都是整数，由于“分形几何”的出现，传统维数的概念已不能概括分形几何中的许多内容。对此我们作以下几点描述。

(1) 自相似理论

自相似概念是分形几何中的重要部分。它的定义是这样，

一种几何图形，如果它可分解成几个部分，而且图形的整体与每个部分都为相似形．由这种相似性可知，对每个局部又可分解成相同个数的相似的子局部。依此类推，一个自相似图形可以有无数多个相似的局部图形组成。

自相似理论有许多现实的应用背景．如地理学中海岸线的图形，化学中某些高分子材料的生长过程及生物医学中的癌细胞的生长过程等．在数学中，典型的自相似图形有如康托(Cantor)集，二维康托集等．康托集的图形如图6.2所示，它把每个线段切去中间的三分之一段．这时由线段$\boldsymbol{G}_0$产生$\boldsymbol{G}_1$，$\boldsymbol{G}_2$，$\boldsymbol{G}_3$，…．对二维康托集也有类似定义，每次切去方块中间的三分之一部分，如图6.3所示。

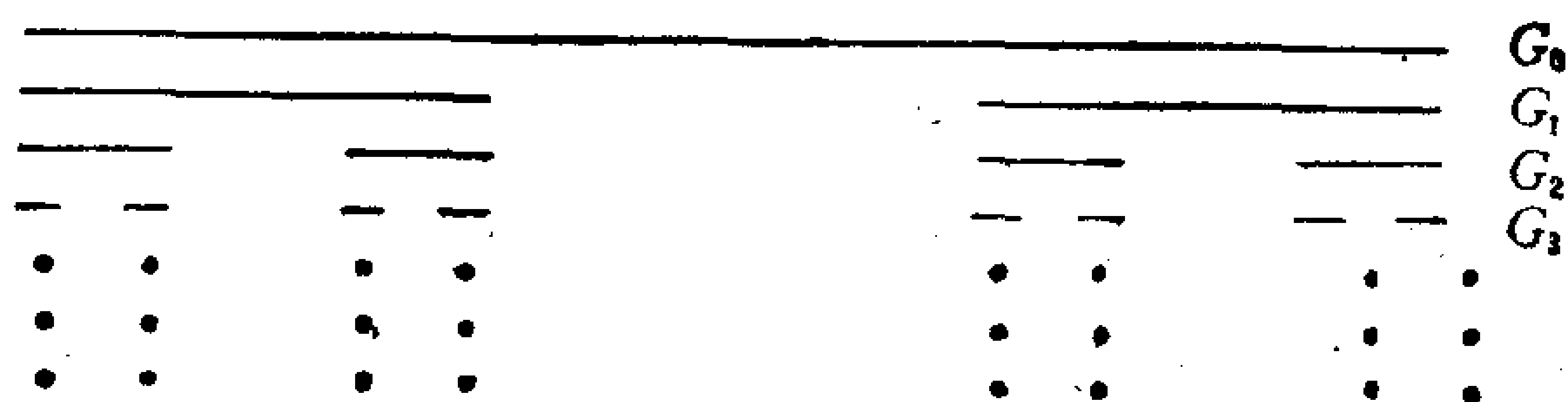


图6.2　康托集

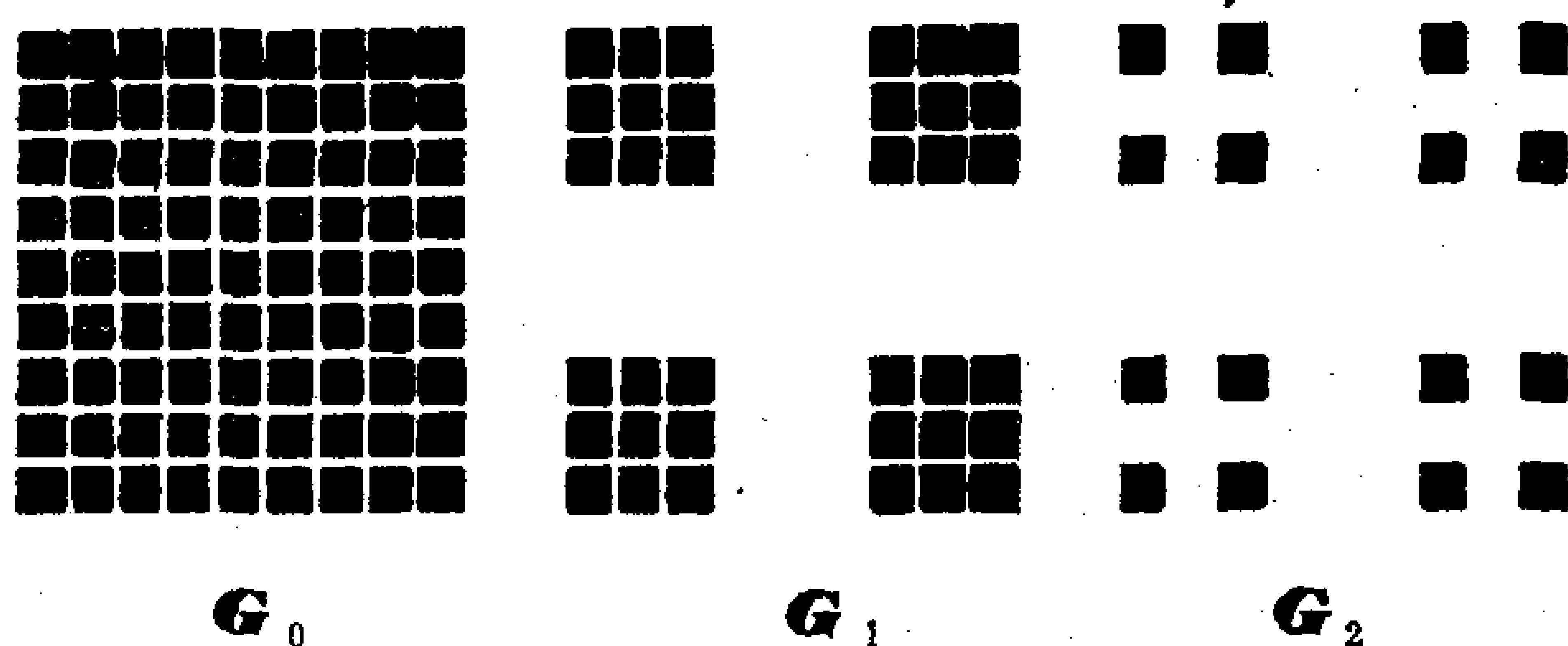


图6.3　二维康托集

(2) 豪斯道夫(Hausdorff)维数

我们现在给出豪斯道夫维数的严格定义．设$\boldsymbol{G}$，$\boldsymbol{U}$是n-维欧氏空间$\boldsymbol{R}^n$中的子集，记$|x-y|$为$\boldsymbol{R}^n$中两点x，y的距离，而

$$|\boldsymbol{U}| \doteq \min\{|x-y|: x,y\in \boldsymbol{U}\} \tag{2.2}$$

为$\boldsymbol{U}$的直径。如果$\boldsymbol{G}\subset\cup_i\boldsymbol{U}_i$，且对每个$i$有$|\boldsymbol{U}_i|\leqslant\delta$，这时称$\{\boldsymbol{U}_i\}$为$\boldsymbol{G}$的一个$\delta$–复盖。

设s为一个非负数，对任何$\delta>0$，我们定义

$$H(\boldsymbol{G},s,\delta)=\inf\{\sum_i|\boldsymbol{U}_i|^s: \{\boldsymbol{U}_i\}\text{为}\boldsymbol{G}\text{的}\delta\text{–复盖}\}, \tag{2.3}$$

$$H(\boldsymbol{G},s)=\lim_{\delta\to 0}H(\boldsymbol{G},s,\ \delta)=\sup_{\delta>0}H(\boldsymbol{G},s,\delta). \tag{2.4}$$

因为$H(\boldsymbol{G},s,\delta)$是$\delta$的单增函数，所以(2.4)式成立。这时称$H(\boldsymbol{G},s)$为$\boldsymbol{G}$的$s$–维豪斯道夫外测度。

定义1 称$s_0=\dim(\boldsymbol{G})$为$\boldsymbol{G}$的豪斯道夫维数，如果s_0满足条件：

$$H(\boldsymbol{G},s)=\begin{cases}\infty, & \text{如果}0\leqslant s<s_0,\\ 0, & \text{如果}s_0<s<\infty.\end{cases} \tag{2.5}$$

由豪斯道夫维数的定义可得康托集$\boldsymbol{G}$的豪斯道夫维数是

$$\dim(\boldsymbol{G})=\log 2/\log 3=0.6309\cdots,$$

且 $$H(\boldsymbol{G})=H(\boldsymbol{G},s_0)=1.$$

3.仙农熵与柯尔莫各洛夫复杂度与豪斯道夫维数的关系

关于柯尔莫各洛夫复杂度与豪斯道夫维数有一系列的专著与论文进行讨论。在本节中我们指出的一点是它们都可与仙农熵建立一系列的等价关系。这些等价条件的讨论文献可在本书最后的索引中找到。

仙农熵与柯尔莫各洛夫复杂度、豪斯道夫维数的等价性的发现在信息论的发展中有十分深远的意义，它不仅使信息论走向信息科学，且使信息的度量向更宽广的领域发展。无论是仙

农熵还是柯尔莫各洛夫复杂度或豪斯道夫维数，它们都是客观世界中一种复杂性的度量。了解这个情形对信息科学今后的发展有重要意义。

§6.3 信息量在统计理论中的应用

信息量在统计理论中的应用范围很多。例如，在谱估计中的熵谱估计理论；假设检验中的误差分析；投影寻踪(Projection Pursuit)中的信息准则及微分几何中的信息统计理论。由于篇幅有限，对这些问题我们不可能作全面介绍。

在本节中，我们要说明的一点就是互熵在微分流形中的作用。如把概率分布空间看作一个黎曼空间，那么费歇(Fisher)信息矩阵就是这个黎曼空间的度规张量，而互熵就是两个概率分布的测地距离。我们认为，这一事实的发现对信息论今后的发展与应用有十分重要的意义。因为，重要空间的距离关系，是该空间理论发展的基础，如欧氏空间中距离关系已成为近代数学的基础支柱之一，而测地距离正是欧氏空间中距离的推广。

另一方面，上述发现把费歇信息、互熵等概念给出了紧密的联系，它们的相互关系不仅得到了统一的解释，而且突出了互熵的地位。这对今后统计理论的深化研究是十分有用的。这样就可利用信息量的理论，得到比常规统计更深入的结果。近几年得到的结果有如：参数估计的高阶逼近；指数型分布中参数的对偶估计；假设检验中两类误差的指数收敛性等。对这些问题的论述涉及许多专门知识与篇幅，有兴趣的读者可参阅有关文献。

结 束 语

在本书中，我们概述了信息度量的类型与应用情况，其中对仙农熵的引入、性质与编码理论作了比较完整的叙述，对无噪声信源编码定理，无记忆信道编码定理；多重信源中的SWC-模型的信源编码定理给出了严格的证明，由此我们可以看到仙农信息论的基本特征与它的经典内容。

由于信息论的近代发展与应用要求，对信息量的定义来源与应用领域日趋扩大。从目前发展的情形来看，这些领域的扩大不仅没有缩小仙农熵的作用，反而增加了它的丰富内涵。许多原来完全不同的概念（如柯尔莫各洛夫复杂度，豪斯道夫维数等）都可与仙农熵发生等价关系。由此可见，信息度量的定义与应用范围及《信息科学》的学科分支正在不断地发展、形成与扩大。

由于篇幅所限，许多有意义的分支与问题在本书中未能提及，如量子信号的信息处理问题；调制编码与光、磁介质材料的信息存储编码；交互信息在投资决策问题中的应用等。这些领域与问题也说明了信息量的广泛应用。

参考文献

〔1〕J, Aczel & Z. Daroczy, On measures of information and their characterization, Academic Press, New York, 1975.

〔2〕S. Amari, Differential-geometrical methods in statistics, Springer–Verlage, New York, 1985.

〔3〕T. Berger, Rate distortion theory: A mathematical basis for data compression, Prentice–Hall, Englewood Cliffs, NJ, 1971.

〔4〕R. E. Blahut, Theory and practice of error control codes, Addison–Wesley, Reading, MA, 1983(徐秉铮等译)。

〔5〕R. E. Blahut, Principles and practice of information theory, Addison–Wesley Reoding MA 1987.

〔6〕T. M. Cover & J. A. Thomas, Elements of information theory, Wiley, New York, 1991.

〔7〕I. Csiszár & J. Körner, Information theory: Coding theorems for discrete memoryless systems, Academic Press, New York, 1981.

〔8〕R. L. Dobrushin, General formulation of Shannon's main theorem of information theory, Usp. Math. Nauk., 14:3–104, 1959. Translated in Am. Math. Soc. Trans., 33, 323–438.

[9] A. El Gamal & T. M. Cover, Multiple user information theory, Proc. IEEE, 68, 1466—1483, 1980.

[10] R. M. Fano, Transmission of information theory: A statistical theory of communication, Wiley, New York, 1961.

[11] K. J. Falconer, The geometry of fractal sets, Cambridge, London, 1986.

[12] A. Feinstein, Foundations of information theory, McGraw-Hill, New York, 1958(江泽培译).

[13] R. G. Gallager, Information theory an dreliable communication, Wiley, New York, 1968.

[14] R. M. Gray, Entropy and information theory, Springer-Verlag, New York, 1990.

[15] C. W. Helstrom, Quantum Detection and estimation theory, Academic, New York, 1976.

[16] Hu Guo Ding, On Shannon theorem and its converse of communication schemes in the case of abstract random variables, Trans. Third Prague Conference on information theory etc., 1962.

[17] 胡国定，信息论中Shennon定理的三种反定理，数学学报，Vol. 11, №2, 260—294, 1961.

[18] 胡国定，沈世镒，Shennon信息论与多用户信息论，电子学报，Vol. 14, No. 434—44, 1986.

[19] A. N. Kolmogorov, Logical basis for information theory and probability theory, IEEE Trans. Inform. Theory, IT—14, 662—664, 1968.

[20] A. G. Konheim, Cryptography A Primer, New York,

1981.
〔21〕 S. Kullback, Information theory and statistics, Wiley, New York, 1959.
〔22〕 S. Kullback, J. C. Keegel & J. H. Kullback, Topics in statistics information theory, Springer-Verlag, New York, 1987.
〔23〕 S. Lin, An introduction to error-correcting codes, Englewood Cliffs, NJ, 1971.
〔24〕 R. J. McEliece, The theory of information, Addiso-Wesley, Reading, MA, 1977.
〔25〕 孟庆生，信息论，西安交大出版社，1987.
〔26〕 M. S. Pinsker, Information and information stability of random variables and processes, Izd. Akad. Nauk, 1960.
〔27〕 B. Ya, Ryabko, Problemy Peredaci Informatsii, 20:3, 16—26, 1986.
〔28〕 C. E. Shannon, A mathematical theory, of communication, Bell Sys. Tech. Journal, 27, 379—423, 623—656, 1948.
〔29〕 沈世镒，平稳通路的基本问题，数学进展，7:1，1—35，1963.
〔30〕 沈世镒，关于《信息论》的若干新方向，数学进展，17:4，351—358，1988.
〔31〕 沈世镒，组合密码学，浙江省科技出版社，1992。
〔32〕 E. C. Van der Meulen, A survey of multi-way channels in information theory, IEEE Trans. Inform. Theory, IT-32, 1—37, 1977.

〔33〕 J. H. Van Lin, Introduction to coding theory, Springer-Verlag, New York, 1982.
〔34〕 万哲先，代数和编码，科学出版社，修订版，1978。
〔35〕 万哲先，戴宗铎，刘木兰，冯绪宁，非线性移位寄存器，科学出版社，1978。
〔36〕 王宏禹，现代谱估计，东南大学出版社，1991。
〔37〕 王育民，何大可，保密学—基础与应用，西北电讯工程学院出版社，1990。
〔38〕 王育民，梁传甲，信息与编码理论，西北电讯工程学院出版社，1986。
〔39〕 J. Wolfowitz, Coding theorems of information theory, Springer-Verlag, Berlin and Prentice-Hill, Englewood Cliffs, NJ, 1978.
〔40〕 杨恩辉，沈世镒，Chaitin复杂度、事件的Shannon信息量与无穷序列 (I), (II)，中国科学A辑，1991:6，591—598；1991:7，710—720。
〔41〕 杨恩辉，Levin's猜想之证明，科学通报，34:21，1761—1765，1989。
〔42〕 章照止，信息量的公理化定义，数学进展，6，289—293，1963。
〔43〕 章照止，信息论与最优编码，上海科技出版社，1992。
〔44〕 章照止，蔡宁，多异度函数的构造和它们的不相容性，系统科学与数学，4，42—54，1984。
〔45〕 周炯磐，信息论基础，人民邮电出版社，1983。

索　引

1.仙农信息论的奠基性工作是〔28〕文，关于仙农熵、信源、信道编码问题乃至率失真理论的原始思想都来自该文。

2.关于仙农熵的引进与基本性质；信源、信道的编码定理在信息论的许多基础教材与专门著作中均有论述，如〔5〕,〔6〕,〔10〕，〔12〕，〔13〕，〔14〕，〔24〕，〔25〕，〔38〕，〔39〕，〔43〕，〔45〕等文．其中，〔12〕文是早期的系统著作，其它各文是在基本问题研究基础上的发展与深化，因此它们各有特色．例如，书〔13〕是一本典型教材；在书〔5〕，〔6〕，〔14〕中，有许多近代研究的成果介绍；〔39〕用组合方法研究编码定理，是国际上近代研究信息论的主流方法；〔43〕是一本基础与专门内容相结合的著作，对国内的研究成果有很多的介绍。

对仙农信息量的特殊补充研究有〔16〕，〔42〕等文。

对编码理论的一般序列模型研究有〔8〕，〔16〕,〔17〕等文，其中〔8〕文用一般概率空间给出了序列模型的一般描述，而〔16〕，〔17〕等文深化研究了序列模的结果，给出了仙农定理的一系列性质。〔16〕文是若干结果的综述。

对有记忆信道的编码问题在〔29〕文中有专门研究。

3.关于连续概率空间中的互熵等理论为专著〔26〕研究，该书是一本对连续概率分布的互熵引进、一般性质与极限理论有全面的论述.而对α-熵、多异度的讨论有〔1〕,〔43〕,〔44〕等文。

对最大熵原理与最小互熵原理，在〔6〕，〔21〕中有完整的介绍，其中〔6〕文给出了各种分布的计算公式.

4.对代数码理论的专著有〔34〕，〔4〕，〔23〕,〔33〕等.另外，在〔10〕，〔24〕，〔25〕等文中也有论述.

对卷积码的讨论有〔5〕，〔6〕，〔24〕，〔25〕，〔43〕等文，其中〔43〕文对卷积码的误差界问题有深入分析.

5.文献〔3〕是研究率失真理论的专著，而〔7〕文是研究多用户信源、信道编码定理的专著. 另外，对多用户信息论的综合介绍论文有〔9〕，〔18〕，〔31〕，〔32〕等文. 在〔5〕，〔6〕，〔24〕，〔43〕等书中也有多用户信息论的讨论.

6.关于密码学的著作有〔20〕，〔31〕，〔35〕，〔37〕等文，另外在〔25〕，〔34〕，〔43〕等书中也有论述.

7.关于柯尔莫各洛夫复杂度的提出是柯尔莫各洛夫的工作，见〔19〕文. 柯尔莫各洛夫复杂度与仙农熵的等价性在书〔6〕中有完整的论述.〔40〕，〔41〕文给出了一系列结果.

文献〔11〕是研究分形几何的专著.〔27〕文给出了关于豪斯道夫维数、柯尔莫各洛夫复杂度与仙农熵的等价性的证明.

8.研究信息与统计的专著有〔2〕，〔21〕，〔22〕等书，其中〔2〕文是在微分流形中研究信息统计问题的著作,对费歇信息，互熵等问题给出了深刻的讨论.

文献〔36〕介绍了近代谱估计的一系列方法，这些方法与最大熵原理等性质相连系.〔15〕文是一本研究量子信号检测与估值的专著，那里的概率分布与仙农熵是希尔伯脱(Hilbert) 空间中的算子.

9.调制编码问题是一种把连续信号直接变为离散信号的编码方式，在近代通信中有重要的应用，对它的处理方式与作用在〔5〕文中有详细的介绍.

交互信息在投资决策问题中有一系列的应用，〔6〕文给出作者在该领域中的一系列工作.